

# Bayesian Compressed Vector Autoregressions\*

**Gary Koop**

University of Strathclyde<sup>†</sup>

**Dimitris Korobilis**

University of Essex<sup>‡</sup>

**Davide Pettenuzzo**

Brandeis University<sup>§</sup>

June 5, 2017

## Abstract

Macroeconomists are increasingly working with large Vector Autoregressions (VARs) where the number of parameters vastly exceeds the number of observations. Existing approaches either involve prior shrinkage or the use of factor methods. In this paper, we develop an alternative based on ideas from the compressed regression literature. It involves randomly compressing the explanatory variables prior to analysis. A huge dimensional problem is thus turned into a much smaller, more computationally tractable one. Bayesian model averaging can be done over various compressions, attaching greater weight to compressions which forecast well. In a macroeconomic application involving up to 129 variables, we find compressed VAR methods to forecast as well or better than either factor methods or large VAR methods involving prior shrinkage.

Keywords: multivariate time series, random projection, forecasting

JEL Classifications: C11, C32, C53

---

\*We would like to thank Andrea Carriero, Todd Clark, Drew Creal, Frank Diebold, Sylvia Kaufmann, Serena Ng, Daniel Peña, Simon Price, Frank Schorfheide, Rob Taylor, Allan Timmermann, Ruey Tsay, Herman van Dijk, Mike West, Jonathan Wright, and Kamil Yilmaz for their helpful comments and suggestions. We also would like to thank seminar participants to the 2016 NBER-NSF Time Series Conference, the 2016 European Seminar on Bayesian Econometrics, the 2016 NBER Summer Institute, the 2016 European Meeting of the Econometric Society, and seminar participants at the Bank of England, Brandeis University, ECARES, University of Essex, University of Konstanz, University of Pennsylvania, and University of St Andrews for their comments.

<sup>†</sup>University of Strathclyde, Department of Economics, Room 4.06, Sir William Duncan Building, 130 Rottenrow, Glasgow G4 0GE, United Kingdom. [gary.koop@strath.ac.uk](mailto:gary.koop@strath.ac.uk)

<sup>‡</sup>Essex Business School, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. [Dimitris.Korobilis@glasgow.ac.uk](mailto:Dimitris.Korobilis@glasgow.ac.uk)

<sup>§</sup>Brandeis University, Sachar International Center, 415 South St, Waltham, MA. [dpettenu@brandeis.edu](mailto:dpettenu@brandeis.edu)

# 1 Introduction

Vector autoregressions (VARs) have been an important tool in macroeconomics since the seminal work of Sims (1980). Recently, many researchers in macroeconomics and finance have been using large VARs involving dozens or hundreds of dependent variables (see, among many others, Banbura, Giannone and Reichlin, 2010, Carriero, Kapetanios and Marcellino, 2009, Koop, 2013, Koop and Korobilis, 2013, Korobilis, 2013, and Gefang, 2014). Such models often have many more parameters than observations, over-fit the data in-sample, and, as a consequence, forecast poorly out-of-sample. Researchers working in the literature typically use prior shrinkage on the parameters to overcome such over-parametrization concerns. The Minnesota prior is particularly popular, but other approaches such as the LASSO (least absolute shrinkage and selection operator, see Park and Casella, 2008 and Gefang, 2014) and SSVS (stochastic search variable selection, see George, Sun and Ni, 2008) have also been used. Most flexible Bayesian priors that result in shrinkage of high-dimensional parameter spaces rely on computationally intensive Markov Chain Monte Carlo (MCMC) methods and their use in recursive forecasting exercises can be computationally infeasible. The only exception is a variant of the Minnesota prior that is based on the natural conjugate prior, an idea that has recently been exploited by Banbura, Giannone and Reichlin (2010) and Giannone, Lenza and Primiceri (2015), among others. While this prior allows for an analytical formula for the posterior, it does have some restrictive features.

The themes of wishing to work with Big Data and needing empirically-sensible shrinkage of some kind also arise in the compressed regression literature; see Donoho (2006). In this literature, shrinkage is achieved by compressing the data instead of the parameters. These methods are used in a variety of models and fields (e.g. neuroimaging, molecular epidemiology, astronomy). A crucial aspect of these methods is that the projections used to compress the data are drawn randomly in a data oblivious manner. That is, the projections do not involve the data and are thus computationally trivial. Recently, Guhaniyogi and Dunson (2015) introduced the idea of Bayesian Compressed regression, where a number of different projections are randomly generated and the explanatory variables are compressed accordingly. Bayesian model averaging (BMA) methods are used to attach different weights to the projections based on the explanatory power the compressed variables have for the dependent variable.

In economics, alternative methods for compressing the data exist. The most popular of

these is principal components (PC) as used, for instance, in the Factor-Augmented VAR, FAVAR, of Bernanke, Boivin and Elias (2005) or the dynamic factor model (DFM) of, e.g., Geweke (1977) and Stock and Watson (2002). PC methods compress the original data into a set of lower-dimensional factors which can then be exploited in a parsimonious econometric specification, for example, a univariate regression or a small VAR. The gains in computation from such an approach are large. However, the data compression is done without reference to the dependent variable(s). PC is thus referred to as an unsupervised data compression method. In contrast, the approach of Guhaniyogi and Dunson (2015) to compressed regression, since it involves the use of BMA, is supervised.

In this paper, we extend the Bayesian random compression methods of Guhaniyogi and Dunson (2015), developed for the regression model, to the VAR case, leading to the Bayesian Compressed VAR (BCVAR). In doing so, we introduce several novel features. First, we generalize the compression schemes of Guhaniyogi and Dunson (2015) and apply them both to the VAR coefficients and the elements of the error covariance matrix. In high dimensional VARs, the error covariance matrix will contain a large number of unknown parameters and, thus, compressing them may be important in avoiding over-parameterization. Second, we allow the explanatory variables in the different equations of the VAR to be compressed in potentially different ways and develop a computationally efficient algorithm that leads to equation-by-equation estimation of the high dimensional compressed VAR.<sup>1</sup> Third, we generalize our compressed VAR methods to the case of large-dimensional VARs with time-varying parameters and volatilities. This model extension is achieved by combining the estimation approach developed in Koop and Korobilis (2013) with the compressed VAR, that is, by relying on variance discounting methods to model, in a computationally efficient way, the time variation in the VAR coefficients and error covariance matrix.

We carry out a substantial macroeconomic forecasting exercise involving VARs with up to 129 dependent variables and 13 lags. We compare the forecasting performance of seven key macroeconomic variables using the BCVAR to various popular alternatives: univariate AR models, the DFM, the FAVAR, and the Minnesota prior VAR. Our results are encouraging for the BCVAR, showing forecast improvements in many cases, and comparable forecast performance in the remainder.

---

<sup>1</sup>This work made use of the High Performance Computing Cluster (HPC64) at Brandeis University. Our algorithm has very low requirements in terms of memory allocation and, since the VAR equations are assumed to be independent, can be easily parallelized to fully exploit the power of modern high-performance computer clusters (HPCC).

## 2 The Theory and Practice of Random Compression

### 2.1 Compression in Regression

Random compression methods have been used in fields such as machine learning and image recognition as a way of projecting the information in data sets with a huge number of variables into a much lower dimensional set of variables. In this way, they are similar to PC methods, which take as inputs many variables and produce factors as their output. With PC methods, the first factor accounts for as much of the variability in the data as possible, the second factor the second most, etc. Typically, a few factors are enough to explain most of the variability in the data and, accordingly, parsimonious models involving only a few factors can be constructed. Random compression does something similar, but is computationally simpler, and capable of dealing with a massively huge number of variables. For instance, in a regression context, Guhaniyogi and Dunson (2015) have an application involving 84,363 explanatory variables.

To fix the basic ideas of random compression, let  $X$  be a  $T \times k$  data matrix involving  $T$  observations on  $k$  variables where  $k \gg T$ .  $X_t$  is a  $1 \times k$  vector denoting the  $t^{th}$  row of  $X$ . Define the projection matrix,  $\Phi$ , which is  $m \times k$  with  $m \ll k$  and  $\tilde{X}_t' = \Phi X_t'$ . Then  $\tilde{X}_t$  is the  $1 \times m$  vector denoting the  $t^{th}$  row of the compressed data matrix,  $\tilde{X}$ . Since  $\tilde{X}$  has  $m$  columns and  $X$  has  $k$ , the former is much smaller and is much easier to work with. To see how this works in a regression context, let  $y_t$  be a scalar dependent variable and consider the relationship:

$$y_t = X_t \beta + \varepsilon_t. \quad (1)$$

If  $k \gg T$ , then working directly with (1) is impossible with some statistical methods (e.g. maximum likelihood estimation) and computationally demanding with others (e.g. Bayesian approaches which require the use of MCMC methods). Some of the computational burden can arise simply due to the need to store in memory huge data matrices. Manipulating such data matrices even a single time can be very demanding. For instance, calculation of the Bayesian posterior mean under a natural conjugate prior requires, among other manipulations, inversion of a  $k \times k$  matrix involving the data. This can be difficult if  $k$  is huge. In order to deal with a large number of predictors, one can specify a compressed regression variant of (1)

$$y_t = (\Phi X_t')' \beta^c + \varepsilon_t. \quad (2)$$

Once the explanatory variables have been compressed (i.e. conditional on  $\Phi$ ), standard

Bayesian regression methods can be used for the regression of  $y_t$  on  $\tilde{X}_t$ . If a natural conjugate prior is used, then analytical formulae exist for the posterior, marginal likelihood, and predictive density, and computation is trivial. Note that the model in (2) has the same structure as a reduced-rank regression, as the  $k$  explanatory variables in the original regression model are squeezed into a small number of explanatory variables given by the vector  $\tilde{X}_t' = \Phi X_t'$ . The crucial difference with previous approaches such as Geweke (1996), Kleibergen and Van Dijk (1998) and Carriero, Kapetanios and Marcellino (2016) is that the matrix  $\Phi$  is not estimated. This is the main idea behind compressed regression methods, where  $\Phi$  is treated as a random matrix with its elements sampled using random number generation schemes.<sup>2</sup>

The key question is: what information is lost by compressing the data in this fashion? The answer is that, under certain conditions, the loss of information may be small. The underlying motivation for random compression arises from the Johnson-Lindenstrauss lemma (see Johnson and Lindenstrauss, 1984). This states that any  $k$  point subset of the Euclidean space can be embedded in  $m = O(\log(k)/\epsilon^2)$  dimensions without distorting the distances between any pair of points by more than a factor of  $1 \pm \epsilon$ , where  $0 < \epsilon < 1$ . In the econometrics literature, Ng (2016, pages 10-13) provides a detailed explanation and the intuition behind this rather remarkable result and shows how it can be used to tackle economic problems. Further intuition on the potential usefulness of these methods in the linear regression setting of (2) can be drawn from the literature on random subspace methods (see Boot and Nibbering, 2016), and complete subset regression (see Elliott, Gargano and Timmermann, 2013, 2015). Both these approaches are similar to the compressed regression in (2). In particular, random subspace methods involve randomly drawing subsets of the explanatory variables, while the complete subset regression method of Elliott, Gargano and Timmermann (2013, 2015) uses equal-weighted combinations of all available subsets of explanatory variables, and resorts to randomly selecting the subsets when the number of regressors is larger than the total number of available observations. Another important reference in this context is Guhaniyogi and Dunson (2015), who provide proofs of the theoretical properties of compressed regression methods, asymptotically in  $T$  and  $k$ . Under some weak assumptions, the most significant relating to sparsity, Guhaniyogi and Dunson (2015) show that their Bayesian compressed regression algorithm produces a predictive density which converges to the true predictive density. The

---

<sup>2</sup>Random projection methods are referred to as *data oblivious*, since  $\Phi$  is drawn without reference to the data. A key early paper in this literature is Achlioptas (2003), which provides theoretical justification for various ways of drawing  $\Phi$  in a computationally-trivial manner.

convergence rate depends on how fast  $m$  and  $k$  grow with  $T$ . With some restrictions on this, they obtain near parametric rates of convergence to the true predictive density. In a simulation study and empirical work, they document excellent coverage properties of predictive intervals and large computational savings relative to popular alternatives. We note that in the large VAR there is likely to be a high degree of sparsity since most VAR coefficients are likely to be zero, especially for more distant lag lengths. In such a case, the theoretical results of Guhaniyogi and Dunson (2015) suggest fast convergence should occur and the computational benefits will likely be large.

Finally, note that Guhaniyogi and Dunson (2015) show that the desirable properties of random compression hold even for a single, data oblivious, random draw of  $\Phi$ . In practice, they recommend taking many random draws and then averaging them. They draw  $\Phi_{ij}$ , the  $ij^{th}$  element of  $\Phi$ , (where  $i = 1, \dots, m$  and  $j = 1, \dots, k$ ) from the following distribution:

$$\begin{aligned} \Pr\left(\Phi_{ij} = \frac{1}{\sqrt{\varphi}}\right) &= \varphi^2 \\ \Pr\left(\Phi_{ij} = 0\right) &= 2(1 - \varphi)\varphi, \\ \Pr\left(\Phi_{ij} = -\frac{1}{\sqrt{\varphi}}\right) &= (1 - \varphi)^2 \end{aligned} \tag{3}$$

where  $\varphi$  and  $m$  are unknown parameters.<sup>3</sup> Next, they rely on BMA to average across the different random projections. Treating each  $\Phi^{(r)}$  ( $r = 1, \dots, R$ ) as defining a new model, they first calculate the marginal likelihood for each model, and then average across the various models using weights proportional to their marginal likelihoods. Note also that  $m$  and  $\varphi$  can be estimated as part of this BMA exercise. In fact, Guhaniyogi and Dunson (2015) recommend simulating  $\varphi$  from the  $U[a, b]$  distribution, where  $a$  ( $b$ ) is set to a number slightly above zero (below one) to ensure numerical stability. As for  $m$ , they recommend simulating it from the discrete  $U[2 \log(k), \min(T, k)]$  distribution.

Intuitively, the use of BMA will ensure that bad compressions (i.e. those that lead to loss of information important for explaining  $y_t$ ) are avoided or down-weighted. To provide some more context, note that if we were to interpret  $m$  and  $\varphi$  and, thus,  $\Phi$ , as random parameters (instead of specification choices defining a particular compressed regression), then BMA can be interpreted as importance sampling. That is, the Uniform distributions that Guhaniyogi and Dunson (2015) use for drawing  $\varphi$  and  $m$  can be interpreted as importance functions. Importance sampling weights are proportional to the posterior for  $m$  and  $\varphi$ . But this, in turn, is equivalent to the marginal likelihood which arises if  $\Phi$  is interpreted as defining a

---

<sup>3</sup>The theory discussed above suggests that  $\Phi$  should be a random matrix whose columns have unit lengths and, hence, Gram-Schmidt orthonormalization is done on the rows of the matrix  $\Phi$ .

model. Thus, in this particular setting, importance sampling is equivalent to BMA. In a VAR context, doing BMA across models should only improve empirical performance since this will lead to more weight being attached to choices of  $\Phi$  which result in superior explanatory power of the compressed data. Such supervised dimension reduction techniques contrast with unsupervised techniques such as PC. It is likely that supervised methods such as this will forecast better than unsupervised methods, a point we investigate in our empirical work.

In summary, for a given compression matrix,  $\Phi$ , the huge dimensional data matrix is compressed into a much lower dimension. This compressed data matrix can then be used in a statistical model such as a regression or a VAR. The theoretical statistical literature on random compression has developed methods such as (3) for randomly drawing the compression matrix and showed them to have desirable properties under weak conditions which are likely to hold in large VARs. By averaging over different  $\Phi$  (which can differ both in terms of  $m$  and  $\varphi$ ) BMA can be done. All this can be done in a computationally simple manner, working only with models of low dimension.

## 2.2 Compression in VARs

To adapt these methods for use with VARs, consider the standard reduced form VAR model,<sup>4</sup>

$$Y_t = BY_{t-1} + \epsilon_t \quad (4)$$

where  $Y_t$  for  $t = 1, \dots, T$  is an  $n \times 1$  vector containing observations on  $n$  time series variables,  $\epsilon_t$  is i.i.d.  $\mathcal{N}(0, \Omega)$  and  $B$  is an  $n \times n$  matrix of coefficients. Note that, with  $n = 100$ , the uncompressed VAR will have 10,000 coefficients in  $B$  and 5,050 in  $\Omega$ . In a VAR(13), such as the one used in this paper, the former number becomes 130,000. It is easy to see why computation can become daunting in large VARs and why there is a need for shrinkage.

To compress the explanatory variables in the VAR, we can use the matrix  $\Phi$  given in (3) but now it will be an  $m \times n$  matrix where  $m \ll n$ , subject to the normalization  $\Phi'\Phi = I$ . In a similar fashion to (2), we can define the compressed VAR:

$$Y_t = B^c(\Phi Y_{t-1}) + \epsilon_t, \quad (5)$$

where  $B^c$  is  $n \times m$ . Thus, we can draw upon the motivations and theorems of, e.g., Guhaniyogi and Dunson (2015) to offer theoretical backing for the compressed VAR. If a natural conjugate

---

<sup>4</sup>For notational simplicity, we explain our methods using a VAR(1) with no deterministic terms. These can be added in a straightforward fashion. In our empirical work, we have monthly data and use 13 lags and an intercept.

prior is used, for a given draw of  $\Phi$  the posterior, marginal likelihood, and predictive density of the compressed VAR in (5) have familiar analytical forms (see, e.g., Koop and Korobilis, 2009). These, along with a method for drawing  $\Phi$ , is all that are required to forecast with the BCVAR. And, if  $m$  is small, the necessary computations of the natural conjugate BCVAR are straightforward.<sup>5</sup> We note however that the natural conjugate prior has some well-known restrictive properties in VARs.<sup>6</sup> In the context of the compressed VAR, working with a  $\Phi$  of dimension  $m \times n$  as defined in (5), with only  $n$  columns instead of  $n^2$  would likely be too restrictive as it implies that lags of all variables are shrunk in the same way in every equation.<sup>7</sup>

An additional issue with the natural conjugate BCVAR is that it allows the error covariance matrix to be unrestricted. This issue does not arise in the regression model of Guhaniyogi and Dunson (2015) but is potentially very important in large VARs. For example, in our application the largest VAR we estimate has an error covariance matrix containing 8,385 unknown parameters. These considerations motivate working with a re-parametrized version of the BCVAR that allows for compression of the error covariance matrix. Following common practice (see, e.g., Primiceri, 2005, Eisenstat, Chan and Strachan, 2015 and Carriero, Clark and Marcellino, 2015) we use a triangular decomposition of  $\Omega$ :

$$A\Omega A' = \Sigma\Sigma, \quad (6)$$

where  $\Sigma$  is a diagonal matrix with diagonal elements  $\sigma_i$  ( $i = 1, \dots, n$ ), and  $A$  is a lower triangular matrix with ones on the main diagonal. Next, we rewrite  $A = I_n + \tilde{A}$ , where  $I_n$  is the  $(n \times n)$  identity matrix and  $\tilde{A}$  is a lower triangular matrix with zeros on the main diagonal. Using this notation, we can rewrite the reduced-form VAR in (4) as follows

$$Y_t = BY_{t-1} + A^{-1}\Sigma E_t$$

where  $E_t \sim N(0, I_n)$ . Further rearranging, we have

$$\begin{aligned} Y_t &= \Gamma Y_{t-1} + \tilde{A}(-Y_t) + \Sigma E_t \\ &= \Theta Z_t + \Sigma E_t \end{aligned} \quad (7)$$

---

<sup>5</sup>In the literature on compression in multivariate regression, it is worth citing Hoff (2007). This paper uses BMA to estimate the rank of a singular value decomposition for the right-hand side variables in a class of models which includes the VAR. In contrast to our approach, he uses Gibbs sampling methods to estimate the optimal decomposition.

<sup>6</sup>These are summarized on pages 279-280 of Koop and Korobilis (2009).

<sup>7</sup>An alternative compressed VAR approach would involve multiplying both sides of the equation by  $\Phi$ , thus compressing the dependent variables as well. In order to forecast, say, the first  $n_f$  variables the upper left hand  $n_f \times n_f$  block of  $\Phi$  could be set to the identity matrix. Such an approach would be similar in spirit to a factor-augmented VAR but with the factors being replaced by random compressions.



where  $Z_t = [Y'_{t-1}, -Y'_t]'$ ,  $\Gamma = AB$  and  $\Theta = [\Gamma, \tilde{A}]$ . Because of the lower triangular structure of  $\tilde{A}$ , the first equation of the VAR above includes only  $Y_{t-1}$  as explanatory variables, the second equation includes  $(Y'_{t-1}, -Y_{1,t})'$ , the third equation includes  $(Y'_{t-1}, -Y_{1,t}, -Y_{2,t})'$ , and so on (here  $Y_{i,t}$  denotes the  $i$ -th element of the vector  $Y_t$ ). Note that this lower triangular structure, along with the diagonality of  $\Sigma$ , means that equation-by-equation estimation of the VAR can be done, a fact we exploit in our algorithm. Furthermore, since the elements of  $\tilde{A}$  control the error covariances, by compressing the model in (7) we can compress the error covariances as well as the reduced form VAR coefficients.

Given that in the triangular specification of the VAR each equation has a different number of explanatory variables, a natural way of applying compression in (7) is through the following specification:

$$Y_{i,t} = \Theta_i^c (\Phi_i Z_t^i) + \sigma_i E_{i,t} \quad i = 1, \dots, n \quad (8)$$

where now  $Z_t^i$  denotes the subset of the vector  $Z_t$  which applies to the  $i$ -th equation of the VAR:  $Z_t^1 = (Y_{t-1})$ ,  $Z_t^2 = (Y'_{t-1}, -Y_{1,t})'$ ,  $Z_t^3 = (Y'_{t-1}, -Y_{1,t}, -Y_{2,t})'$ , and so on. Similarly,  $\Phi_i$  is a matrix with  $m$  rows and column dimension that conforms with  $Z_t^i$ . Following (8), we now have  $n$  compression matrices (each of potentially different dimension and with different randomly drawn elements), and as a result the explanatory variables in the equations of the original VAR can be compressed in different ways.<sup>8</sup>

For a given set of posterior draws of  $\Theta_i^c$  and  $\sigma_i$  ( $i = 1, \dots, n$ ), estimation and prediction can be done in a computationally-fast fashion using a variety of methods since each model will be of low dimension and, for the reasons discussed previously, all these can be done one equation at a time. In the empirical work in this paper, we use standard Bayesian methods suggested in Zellner (1971) for the seemingly unrelated regressions model. In particular, for each equation we use the prior:

$$\begin{aligned} \Theta_i^c | \sigma_i^2 &\sim N(\underline{\Theta}_i^c, \sigma_i^2 \underline{V}_i) \\ \sigma_i^{-2} &\sim G(\underline{s}_i^{-2}, \underline{\nu}_i), \end{aligned} \quad (9)$$

---

<sup>8</sup>Note also that an alternative way to estimate a compressed VAR version of model (7) would be to write the model in its SUR form; see Koop and Korobilis (2009). If we did so, the data matrix  $Z_t$  would have to be expanded by taking its Kronecker product with  $I_n$ . For large  $n$  such an approach would require huge amounts of memory (many times more than a modern personal computer has available). Even if we were to use sparse matrix calculations, having to define the non-zero elements of the matrices in the SUR form of a large VAR would result in very slow computation. On the other hand, the equation-by-equation estimation we propose in (8) is simpler and can be easily parallelizable, since the VAR equations are transformed so as to be independent.

where  $G(\underline{s}_i^{-2}, \underline{\nu}_i)$  denotes the Gamma distribution with mean  $\underline{s}_i^{-2}$  and degrees of freedom  $\underline{\nu}_i$ . In our empirical work, we set  $\underline{\Theta}_i^c = 0$ ,  $\underline{V}_i = 0.5 \times I$  and, for  $\sigma_i^{-2}$  use the non-informative version of the prior (i.e.  $\underline{\nu}_i = 0$ ). We then use familiar Bayesian results for the Normal linear regression model (e.g. Koop, 2003, page 37) to obtain analytical posteriors for both  $\Theta_i^c$  and  $\sigma_i^2$ . The one-step ahead predictive density is also available analytically. However,  $h$ -step ahead predictive densities for  $h > 1$  are not available analytically. To compute them, we proceed by first converting the estimated compressed triangular VAR in equation (8) back into the triangular VAR of equation (7), noting that

$$\Theta = [(\Theta_1^c \Phi_1, 0_n)', (\Theta_2^c \Phi_2, 0_{n-1})', \dots, (\Theta_{n-1}^c \Phi_{n-1}, 0_2)', (\Theta_n^c \Phi_n, 0)']' \quad (10)$$

where  $0_n$  is a  $(1 \times n)$  vector of zeros,  $0_{n-1}$  is a  $(1 \times n - 1)$  vector of zeros, and so on. Subsequently, we go from the triangular VAR in equation (7) to the original reduced-form VAR in equation (4) by noting that  $B = A^{-1}\Gamma$ , where  $\Gamma$  can be recovered from the first  $n \times n$  block of  $\Theta$  in (10), and  $A$  is constructed from  $\tilde{A}$  using the remaining elements of  $\Theta$  (see equation (7)). Finally, the covariance matrix of the reduced form VAR is simply given by equation (6), where both  $A$  and  $\Sigma$  are known. After these transformations are implemented, standard results for Bayesian VARs can be used to obtain multi-step-ahead density forecasts.

So far we have discussed specification and estimation of the compressed VAR conditional on a single compression  $\Phi$  (or  $\Phi_i$ ,  $i = 1, \dots, n$ ). In practice, we generate  $R$  sets of such compression matrices  $\Phi_i^{(r)}$  ( $i = 1, \dots, n$  and  $r = 1, \dots, R$ ), and estimate an equal number of compressed VAR models, which we denote with  $M_1, \dots, M_R$ . Then, for each model, we use the predictive simulation methods described above to obtain the full predictive density  $p(Y_{t+h}|M_r, \mathcal{D}^t)$ , where  $h = 1, \dots, H$ . For each forecast horizon  $h$ , the final BMA forecast is a mixture of the form

$$p(Y_{t+h}|\mathcal{D}^t) = \sum_{r=1}^R w_r p(Y_{t+h}|M_r, \mathcal{D}^t), \quad (11)$$

where  $\mathcal{D}^t$  is the information set available at time  $t$ ,  $w_r = \exp(-.5\Psi_r) / \sum_{r=1}^R \exp(-.5\Psi_r)$  is model  $M_r$  weight, and  $\Psi_r = BIC_r - BIC_{\min}$ , with  $BIC_r$  being the value of the Bayesian Information Criterion (BIC) of model  $M_r$  and  $BIC_{\min}$  the minimum value of the BIC among all  $R$  models. We use BIC to approximate the marginal likelihood because it can be computed easily for high-dimensional VARs and is insensitive to the choice of the priors.

In our empirical work, the  $\Phi_i^{(r)}$ 's are randomly drawn using the strategy described in (3). This scheme means that for each of the  $R$  random compression matrices, we have to generate

the parameter  $\varphi$  and decide on the number of rows  $m$  of each  $\Phi_i^{(r)}$  (that is, the dimension of the projected space). Both these parameters are drawn randomly:  $\varphi$  is drawn from the uniform  $U[0.1, 0.8]$  distribution and  $m$  is drawn from the discrete  $U[1, 5 \ln(k_i)]$ , where  $k_i$  is the number of explanatory variables included in  $Z_t^i$  for VAR equation  $i$ .<sup>9</sup>

We note that papers such as Achlioptas (2003) have proposed alternative schemes to the one we adopted in (3) to randomly draw the elements of  $\Phi_i$ . While some of these may be potentially more efficient and can provide a higher degree of sparsity (zeros in  $\Phi_i$ ), in our macroeconomic application we found that a wide range of alternative random projection schemes produced almost identical forecasts. Thus, in our empirical application we will focus exclusively on the scheme proposed by Guhaniyogi and Dunson (2015), as described in equation (3).

### 2.3 Compression in Time-Varying Parameter VARs

In macroeconomic forecasting applications, it is often empirically necessary to allow for time-variation in the VAR coefficients and/or the error covariance matrix. There is an increasing literature that shows that ignoring macroeconomic volatility and possible structural changes in coefficients of a VAR can result in bad in-sample fit and poor out-of-sample forecast performance; see for example Clark (2011). Both such extensions add greatly to the computational burden since MCMC methods are usually required. In the context of the constant coefficient VAR with conjugate prior for the VAR coefficients there is a growing literature (e.g. Carriero, Clark and Marcellino, 2015, 2016a and Chan, 2015) investigating various structures for time-varying error covariance matrices which do not lead to excessively large computational demands. However, even these can be restrictive and require the use of MCMC methods which will make them unsuitable for use in extremely large models. Allowing for time-variation in the VAR coefficients (e.g. assuming that the coefficients evolve according to a random walk or a Markov switching process) will also greatly increase the burden.

In this section, we show how the compressed VAR methods can be generalized to the case of a VAR with time-varying parameters and stochastic volatilities (TVP-SV-VAR). We will

---

<sup>9</sup>These choices are similar to those used in Guhaniyogi and Dunson (2015), but our choice to draw values of  $m$  as low as one is lower than theirs. We do this just to see if extreme compressions, which basically remove all the right-hand side variables, receive any support. Due to numerical stability reasons, for  $\varphi$  we do not consider the full support  $[0, 1]$ .

denote our compressed version of the TVP-SV-VAR as  $\text{BCVAR}_{tvp-sv}$  and write it as:

$$Y_{i,t} = \Theta_{i,t}^c (\Phi_i Z_t^i) + \sigma_{i,t} E_{i,t}. \quad (12)$$

Notice that all parameters including the error variances may vary over time and, thus, they have  $t$  subscripts,  $t = 1, \dots, T$ . We also remind the reader that the variables  $Z_t^i$  contain lags of the dependent variables and the terms which relate to the error covariances as defined in (7). This TVP-SV-VAR model is different from the previous literature because it allows for equation by equation estimation. Papers such as Primiceri (2005) would specify the VAR in the familiar seemingly unrelated regression form, where all  $n$  VAR equations are modeled jointly. Estimation using the latter form can become cumbersome as  $n$  increases, since the posterior for both the time-varying regression coefficients and volatilities requires many manipulations involving large data matrices. Using (12), estimation of the  $\text{BCVAR}_{tvp-sv}$  is reduced to the estimation of  $n$  univariate time-varying parameter regressions which is computationally more efficient for large  $n$ . Additionally, the possibly large matrix  $Z_t^i$  is still compressed using  $\Phi_i$  as with the BCVAR.

In general, forecasting with TVP-SV-VARs is computationally demanding as it typically relies on MCMC methods. In our case, even if we use  $\Phi_i$  to compress the data, a full Bayesian analysis could be computationally demanding with large  $n$  since MCMC methods are required and must be run for each of the  $n$  equations. Accordingly, we turn to approximate methods to deal with the TVP-SV aspect of our BCVAR. These are generalizations of those developed by Koop and Korobilis (2013) in the context of a TVP-VAR with time varying error covariance matrix. They use variance discounting methods to model the time-variation in the VAR coefficients and error covariance matrix, and provide analytical formulae for updating them. Thus, in (12), once we draw  $\Phi_i$  randomly,  $\Theta_{i,t}^c$  and  $\sigma_{i,t}^2$  can be updated using simple recursive formulae based on the Kalman filter, without relying on computationally intensive MCMC methods.

Adapting Koop and Korobilis (2013), the compressed TVP-SV-VAR model involves estimating  $\Theta_{i,t}^c$  and  $\sigma_{i,t}^2$  by assuming that they evolve according to:

$$\Theta_{i,t}^c = \Theta_{i,t-1}^c + \sqrt{\frac{(1 - \lambda_{i,t}) \text{var}(\Theta_{i,t-1}^c | t-1)}{\lambda_{i,t}}} u_{i,t}, \quad (13)$$

$$\sigma_{i,t}^2 = \kappa_{i,t} \sigma_{i,t-1}^2 + (1 - \kappa_{i,t}) \hat{E}_{i,t}^2. \quad (14)$$

That is,  $\Theta_{i,t}^c$  follows a random walk using a forgetting factor approximation to its error covariance matrix. Kalman filtering methods can be used for this equation. For  $\sigma_{i,t}^2$  we have

an Exponentially Weighted Moving Average filter.  $\hat{E}_{i,t}^2$  is the time  $t$  squared prediction error estimated from the  $i$ -th equation of the VAR,  $u_{i,t} \sim N(0, 1)$ , and  $\text{var}(\Theta_{i,t-1}^c)$  is the variance of  $\Theta_{i,t-1}^c$  given information up to time  $t-1$  and is produced by the Kalman filter (see Koop and Korobilis, 2013, for details). The crucial parameters in this specification are the forgetting and decay factors  $\lambda_{i,t}$  and  $\kappa_{i,t}$ . These factors, which are typically in the range of  $(0.9, 1)$ , control how quickly discounting of past data occurs. For example, if  $\lambda_{i,t} = 0.90$  then  $\Theta_{i,t}^c$  depends very heavily on recent observations, and changes very rapidly over time. On the other hand, if  $\lambda_{i,t} = 0.99$  the discounting of the past is more gradual and  $\Theta_{i,t}^c$  varies more smoothly. Finally, when  $\lambda_{i,t} = 1$  we go back to the constant parameter VAR. Similar arguments can be made for  $\sigma_{i,t}^2$  and its decay factor  $\kappa_{i,t}$ .

We extend the methods of Koop and Korobilis (2013) by allowing for the decay and forgetting factors to vary over time using simple updating formulae:

$$\lambda_{i,t} = \underline{\lambda} + (1 - \underline{\lambda}) \times \exp\left(-0.5 \times \frac{\hat{E}_{i,t-1}^2}{\hat{\sigma}_{i,t-1}^2}\right), \quad (15)$$

$$\kappa_{i,t} = \underline{\kappa} + (1 - \underline{\kappa}) \times \exp\left(-0.5 \times \text{kurt}\left(\hat{E}_{i,t-12:t-1}\right)\right), \quad (16)$$

where  $\hat{\sigma}_{i,t-1}^2$  is the time  $t-1$  estimate of the variance and  $\text{kurt}\left(\hat{E}_{i,t-12:t-1}\right)$  is the excess kurtosis of the VAR prediction error, evaluated over the past year (i.e. with monthly data this is based on a rolling sample of 12 observations).  $\underline{\lambda}$  and  $\underline{\kappa}$  put bounds on the minimum values of the forgetting and decay factors. We set  $\underline{\lambda} = 0.98$  and  $\underline{\kappa} = 0.94$  which, in the context of monthly data, allow for the possibility of a fairly large amount of time variation.<sup>10</sup>

Note that if the prediction error is close to zero then  $\lambda_{i,t} = 1$ , which is the value consistent with the parameters in equation  $i$  being constant. In words, if the model forecast well last month, we do not change its parameters this month. However, the larger the prediction error is, the smaller  $\lambda_{i,t}$  becomes and, thus, a higher degree of parameter change is allowed for. For the decay factor  $\kappa_{i,t}$ , we use a similar reasoning, except for the fact that we do so in terms of the excess kurtosis of the prediction error. As is well known (e.g. from the GARCH literature), under the assumption that errors are Normally distributed, in times of constant volatility the excess kurtosis will be equal to zero, while in times of increased volatility the excess kurtosis will be higher. Allowing for  $\kappa_{i,t}$  to depend on the kurtosis over the past year is a simple way of allowing  $\sigma_{i,t}^2$  to change more rapidly in unstable times. Using these

---

<sup>10</sup>The idea of allowing the value of the forgetting factor to depend on the most recent prediction error is used, e.g., in Park, Jun, and Kim (1991).

methods, it is straightforward to allow for time-variation in our compressed VAR approach in a computationally simple manner.

### 3 Empirical Application: Macroeconomic Forecasting with Large VARs

#### 3.1 Data

We use the FRED-MD data-base of monthly US variables from January 1960 through December 2014. The reader is referred to McCracken and Ng (2015) for a description of this macroeconomic data set, which includes several variables from a broad range of categories (e.g. output, capacity, employment and unemployment, prices, wages, housing, inventories and orders, stock prices, interest rates, exchange rates and monetary aggregates). We use the 129 variables for which complete data is available, after transforming all variables using the transformation codes provided in the online appendix.<sup>11</sup> We present forecasting results for seven variables of interest: industrial production growth (INDPRO), the unemployment rate (UNRATE), total nonfarm employment (PAYEMS), the change in the Fed funds rate (FEDFUNDS), the change in the 10 year T-bill rate (GS10), the finished good producer price inflation (PPIFGS) and consumer price inflation (CPIAUCSL).<sup>12</sup> In particular, we estimate VARs of different dimensions, with these seven variables included in all of our specifications. We have a Medium VAR with 19 variables and a Large VAR with all 129 variables.<sup>13</sup> A listing of all variables, including exact details of which variables appear in which VAR, is given in the online appendix. Note that most of our variables have substantial persistence in them and, as a result, the first own lag in each equation almost always has important explanatory power. Accordingly, we do not compress the first own lag. This is included in every equation, with compression being done on the remaining variables.<sup>14</sup> Following Banbura et al. (2010), we choose a relatively large value for the lag length ( $p = 13$ ) for all the methods we compare, trusting in the compression or shrinkage of

---

<sup>11</sup>In addition to dropping a few series with missing observations, we also remove the series non-borrowed reserves, as it became extremely volatile during the Great Recession.

<sup>12</sup>We also standardize our variables prior to estimation and forecasting. The forecasts of the original variables are then computed by inverting the transformation. This standardization is computed recursively, i.e., using only the data that would have been available at each point in time to estimate the various models.

<sup>13</sup>In our online appendix, we also present results for an Intermediate VAR with 46 variables.

<sup>14</sup>To be precise, we are always allowing the diagonal elements of  $\Gamma$  in (7) to be non-zero. We experimented with the alternative triangularization of Carriero, Clark and Marcellino (2016b) which allows for the diagonal elements of  $B$  in (4) to always be non-zero. These two approaches yield very similar results, both in terms of treatment of first own lags and in forecast performance.

the various methods to remove unnecessary lags.

### 3.2 Alternative Methods for Large VARs

We use the Bayesian compressed VAR methods introduced in [subsection 2.2](#) in two ways: the first one, which we label as  $\text{BCVAR}_c$ , compresses both the VAR coefficients and the error covariances as in (8). The second one, which we label  $\text{BCVAR}$ , is the same, except for the fact that it does not compress the error covariances.

To better assess the forecasting accuracy of these compressed VAR methods, we compare their performances against a number of popular alternatives. Reasoning that previous work with large numbers of dependent variables have typically used factor methods or large Bayesian VARs, we focus on these. In addition, we compare the forecasts using all of these methods to a benchmark approach which uses OLS forecasts from univariate  $\text{AR}(1)$  models.

#### Dynamic Factor Model

The dynamic factor model (DFM) can be written as:

$$\begin{aligned} Y_t &= \lambda_0 + \lambda_1 F_t + \epsilon_t \\ F_t &= \Phi_1 F_{t-1} + \dots + \Phi_p F_{t-p} + \epsilon_t^F \end{aligned} \tag{17}$$

where  $F_t$  is a  $q \times 1$  vector of factors (with  $q \ll n$ ) which contains information extracted from all  $n$  variables,  $\lambda_0$  and  $\lambda_1$  are  $n \times 1$  and  $n \times q$  matrices, and  $\epsilon_t \sim \mathcal{N}(0, \Sigma^Y)$  where  $\Sigma^Y$  is a diagonal matrix. The vector of factors is assumed to follow a  $\text{VAR}(p)$  process with  $\epsilon_t^F \sim \mathcal{N}(0, \Sigma^F)$ , with  $\epsilon_t$  independent of  $\epsilon_s^F$  at all  $t$  and  $s$ . We use principal component methods to estimate the factors.<sup>15</sup>

We specify the maximum number of factors and lag lengths to be  $q^{\max} = \sqrt{n}$  and  $p^{\max} = 13$ , respectively. Next, at each point in time we use  $BIC$  to choose the optimal lag length and number of factors. We use Bayesian methods with non-informative priors to estimate and forecast with this model (note that the law of motion for the common factors in equation (17) is needed to iterate forward the forecasts when  $h > 1$ ).

#### Factor-Augmented VAR

We use the Factor-Augmented VAR (FAVAR) of Bernanke, Boivin, Elias (2005) dividing  $Y_t$  into a set of primary variables of interest,  $Y_t^*$  (these are the same key seven variables listed

---

<sup>15</sup>Alternative estimators such as the quasi-maximum likelihood estimator of Doz, Giannone and Reichlin (2012) are possible. These authors note that principal components, quasi maximum likelihood, and a two-step estimator based on Kalman smoother all give basically the same results for  $n > 25$  and  $T > 50$ . We use principal components for simplicity.

above), and the remainder  $\tilde{Y}_t$ , and work with the model:

$$\begin{aligned}\tilde{Y}_t &= \Lambda F_t + \epsilon_t^{\tilde{Y}} \\ \begin{bmatrix} F_t \\ Y_t^* \end{bmatrix} &= B_0 + B_1 \begin{bmatrix} F_{t-1} \\ Y_{t-1}^* \end{bmatrix} + \dots + B_p \begin{bmatrix} F_{t-p} \\ Y_{t-p}^* \end{bmatrix} + \epsilon_t^*.\end{aligned}\tag{18}$$

The vector  $(F_t', Y_t^{*'})'$  is assumed to follow a VAR( $p$ ) process with  $\epsilon_t^{\tilde{Y}} \sim \mathcal{N}(0, \Sigma^{\tilde{Y}})$ ,  $\epsilon_t^* \sim \mathcal{N}(0, \Sigma^*)$ ,  $E(\epsilon_t^{\tilde{Y}} \epsilon_s^{*'}) = 0$  for all  $t$  and  $s$  and  $E(\epsilon_t^* \epsilon_s^{*'}) = E(\epsilon_t^{\tilde{Y}} \epsilon_s^{\tilde{Y}'}) = 0$  for all  $s \neq t$ . As with the DFM model, we rely on principal component methods to extract the factors  $F_t$ , and select the optimal number of factors  $q$  and the lag length  $p$  using *BIC*. We use Bayesian methods with non-informative priors to forecast with this model.

### Bayesian VAR using the Minnesota Prior

We follow closely Banbura et al (2010)'s implementation of the Minnesota prior VAR which involves a single prior shrinkage parameter,  $\omega$ . However, we select  $\omega$  in a different manner than Banbura et al (2010), and estimate it in a data-based fashion similar to Giannone, Lenza and Primiceri (2015). We choose a grid of values for the inverse of the shrinkage factor  $\omega^{-1}$  ranging from  $0.5 \times \sqrt{np}$  to  $10 \times \sqrt{np}$ , in increments of  $0.1 \times \sqrt{np}$ . At each point in time, we use *BIC* to choose the optimal degree of shrinkage. All remaining specification and forecasting choices are exactly the same as in Banbura et al (2010) and, hence, are not reported here. In our empirical results, we use the acronym BVAR to refer to this approach.

We stress that we are only comparing our methods to alternatives that are computationally feasible with large VARs. This restriction rules out many popular VAR-based approaches and explains why we are only considering the Minnesota prior VAR. But we note that even the Minnesota prior VAR will not handle the truly enormous VARs that may arise for the researcher working with multi-country data sets or combining macroeconomic and financial data. In contrast, random compression methods should scale up to handle VARs with thousands of variables (as will principal components methods). Carriero, Clark and Marcellino (2016a) explore in detail the computational challenges of working with large VARs and note that the posterior covariance matrix for the VAR is an  $(np + 1)n$  matrix whose manipulation can cause a chief computational bottleneck. With general approaches (which do not involve a natural conjugate or Minnesota prior) manipulating such a matrix involves  $O(n^6 p^3)$  operations, but with priors adopting a particular Kronecker structure (e.g. the Minnesota prior) this can be reduced to  $O(n^3 p^3)$ . When  $n = 100$  or larger this results in a huge computational reduction, which is why so many large VAR applications rely on priors which have this Kronecker structure (despite



well known criticisms of it). But when  $n = 1,000$  or  $n = 10,000$  even if the Kronecker structure is maintained there will come a point where computation will break down. Furthermore, when forecasting with large VARs the Minnesota prior is mainly used for point forecasting since obtaining the predictive density typically involves Monte Carlo predictive simulation, that is, simulation of VAR coefficients followed by simulation of future values of the dependent variables. The need for simulation procedures raises additional computational bottlenecks which limit the use of the Minnesota prior VAR in very high dimensions.

### 3.3 Measures of Predictive Accuracy

We use the first half of the sample, January 1960–June 1987, to obtain initial parameter estimates for all models, which are then used to predict outcomes from July 1987 ( $h = 1$ ) to June 1988 ( $h = 12$ ). The next period, we include data for July 1987 in the estimation sample, and use the resulting estimates to predict the outcomes from August 1987 to July 1988. We proceed recursively in this fashion until December 2014, thus generating a time series of forecasts for each forecast horizon  $h$ , with  $h = 1, \dots, 12$ . Note that when  $h > 1$ , point forecasts are iterated and predictive simulation is used to produce the predictive densities.

Next, for each of the seven key variables listed above we summarize the precision of the  $h$ -step-ahead point forecasts for model  $i$ , relative to that from the univariate AR(1), by means of the ratio of MSFEs:

$$MSFE_{ijh} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} e_{i,j,\tau+h}^2}{\sum_{\tau=\underline{t}}^{\bar{t}-h} e_{bck,j,\tau+h}^2}, \quad (19)$$

where  $\underline{t}$  and  $\bar{t}$  denote the start and end of the out-of-sample period, and where  $e_{i,j,\tau+h}^2$  and  $e_{bck,j,\tau+h}^2$  are the squared forecast errors of variable  $j$  at time  $\tau$  and forecast horizon  $h$  associated with model  $i$  ( $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ ) and the AR(1) model, respectively. The point forecasts used to compute the forecast errors are obtained by averaging over the draws from the various models'  $h$ -step-ahead predictive densities. Values of  $MSFE_{ijh}$  below one suggest that model  $i$  produces more accurate point forecasts than the AR(1) benchmark for variable  $j$  and forecast horizon  $h$ .

We also assess the accuracy of the point forecasts of the various methods using the multivariate loss function of Christoffersen and Diebold (1998). Specifically, we compute the ratio between the multivariate weighted mean squared forecast error (WMSFE) of model  $i$

and the WMSFE of the benchmark AR(1) model as follows:

$$WMSFE_{ih} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{i,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{bcmk,\tau+h}}, \quad (20)$$

where  $we_{i,\tau+h} = (e'_{i,\tau+h} \times W \times e_{i,\tau+h})$  and  $we_{bcmk,\tau+h} = (e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h})$  are time  $\tau+h$  weighted forecast errors of model  $i$  and the benchmark model,  $e_{i,\tau+h}$  and  $e_{bcmk,\tau+h}$  are the  $(7 \times 1)$  vector of forecast errors for the key series we focus on, and  $W$  is a  $(7 \times 7)$  matrix of weights. We set the matrix  $W$  to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast.

As for the quality of the density forecasts, we follow Geweke and Amisano (2010) and compute the average log predictive likelihood differential between model  $i$  and the AR(1) benchmark,

$$ALPL_{ijh} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} (LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h}), \quad (21)$$

where  $LPL_{i,j,\tau+h}$  ( $LPL_{bcmk,j,\tau+h}$ ) denotes model  $i$ 's (benchmark's) log predictive score of variable  $j$ , computed at time  $\tau+h$ , i.e., the log of the  $h$ -step-ahead predictive density evaluated at the outcome. Positive values of  $ALPL_{ijh}$  indicate that for variable  $j$  and forecast horizon  $h$ , model  $i$  produces on average more accurate density forecasts than the benchmark model.

Finally, we consider the multivariate average log predictive likelihood differentials between model  $i$  and the benchmark AR(1),

$$MVALPL_{ih} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} (MVLPL_{i,\tau+h} - MVLPL_{bcmk,\tau+h}), \quad (22)$$

where  $MVLPL_{i,\tau+h}$  and  $MVLPL_{bcmk,\tau+h}$  denote the multivariate log predictive likelihoods of model  $i$  and the benchmark model at time  $\tau+h$ , computed under the assumption of joint normality.

In order to test the statistical significance of differences in point and density forecasts, we consider pairwise tests of equal predictive accuracy (henceforth, EPA; Diebold and Mariano, 1995; West, 1996) in terms of MSFE, WMSFE, ALPL, and MVALPL. All EPA tests we conduct are based on a two sided test with the null hypothesis being the AR(1) benchmark, and standard normal critical values. Based on simulation evidence in Clark and McCracken (2013), when computing the variance estimator which enters the test statistic we rely on serial correlation robust standard errors, and incorporate a finite sample correction (Harvey et al.,

1997). In the tables, we use \*\*\*, \*\* and \* to denote results which are significant at the 1%, 5% and 10% levels, respectively, in favor of the model listed at the top of each column.

### 3.4 Forecasting Results for VAR and Factor Model Approaches

We begin by considering all our models with constant coefficients. Tables 1, 3, and the left side of Table 5 present evidence on the quality of our point forecasts for the seven main variables of interest, relative to the AR(1) benchmark. We are finding that in the majority of cases BCVAR methods beat the benchmark. Additionally, they often tend to forecast better than the other approaches. Table 5, which presents the WMSFEs over the seven variables of interest, provides the best overall summary of our results as they relate to point forecasts. With six forecast horizons and two VAR sizes, this table contains 12 dimensions in which point forecasts can be compared. In 11 of these, either BCVAR or BCVAR<sub>c</sub> is the model with the lowest WMSFE. In six of these cases, compressed VAR approaches beat the benchmark in a statistically significant manner. The FAVAR is the next best approach, although it is worth noting that in some cases (e.g. with short term forecasting and particularly with the Medium VAR) it does poorly, failing to beat the AR(1) benchmark. Overall, our results indicate that random compression works quite well, often producing the best forecasts. In addition, in those instances when that is not the case it appears that random compression still works quite well. This result suggests that a risk averse user might feel confident using random compression methods. In summary, random compression of the VAR coefficients is at least competitive with other multivariate forecasting methods with the data set under consideration. Evidence relating to compression of the error covariance is more mixed. That is, in some instances the BCVAR<sub>c</sub> forecasts better than the BCVAR, but there are many cases where the forecasts from the BCVAR model are more accurate.

With regards to forecast horizon, no clear pattern emerges. There is a slight tendency for compressed VAR approaches to do particularly well at shorter horizons, but there are no strong differences across horizons. In terms of the individual variables, one notable pattern in these tables is that BCVAR and BCVAR<sub>c</sub> are (with some exceptions) forecasting particularly well for the most important macroeconomic aggregates such as prices, unemployment and industrial production. In contrast, for the long-term interest rate (GS10), our Large VAR is almost never beating the benchmark. But even in this case, where small models are forecasting well, it is reassuring to see that the MSFEs obtained using random compression methods are only slightly worse than the benchmark ones. This result indicates that random compression

methods are finding that the GS10 equation in the Large VAR is hugely over-parametrized, and are successfully compressing the explanatory variables in a way to obtain results that are nearly as good as those from the more parsimonious univariate models.

Figures 1 and 2 present evidence on when the forecasting gains of BCVARs, relative to the other approaches, are achieved. These figures plot the cumulative sum of weighted forecasting errors (jointly for the seven variables of interest) for the benchmark AR(1) model minus those from a competing approach,  $CSWFED_{iht} = \sum_{\tau=t}^{t-h} (we_{bcmk, \tau+h} - we_{i, \tau+h})$ , for different VAR sizes and different forecasting horizons. Positive values of this metric imply that an approach is beating the benchmark. For short horizons, BCVAR is the only approach that consistently beats the benchmark model, throughout the whole forecast period. All other approaches accumulate more forecast errors over time compared to the simple AR(1). It is also interesting to note that during the 2007-2009 crisis all multivariate methods seem to, at least temporarily, improve over the univariate AR(1). However, towards the end of the crisis, for all methods but the BCVAR, relative forecast performance deteriorates abruptly. As for the longer forecast horizons, some of the alternative multivariate models seem to perform fairly well. This is especially true for the FAVAR model, which for the Medium VAR and  $h = 12$  ends up being the best model.<sup>16</sup>

Tables 2, 4, and the right hand side of Table 5 shed light on the quality of our density forecasts, by presenting averages of log predictive likelihoods for the VARs of different dimensions. Results for the ALPLs appear qualitatively similar to those for the MSFEs, so we will not discuss them in detail. But they do differ in their strength in two ways. First, the evidence that compressed VAR approaches can beat univariate benchmarks becomes stronger. See in particular the right hand side of Table 5, which shows strong rejections of the hypothesis of EPA at every horizon and for every VAR dimension. Second, the evidence that compressed VARs can forecast better than BVAR or FAVAR approaches becomes somewhat weaker. In particular, for the Medium VAR standard Bayesian VAR methods based on the Minnesota prior tend to forecast slightly better than the compressed VAR approaches. Nevertheless, our BCVAR does particularly well in the Large VAR case, improving over the standard large Bayesian VAR and FAVAR methods at all forecast horizons.

Figures 3 and 4 plot the cumulative sums of the multivariate log predictive likelihood

---

<sup>16</sup>Additional results, including plots of cumulative sum of squared forecast errors for the individual variables, are available in the online appendix.

differentials,  $CSMVLPD_{iht} = \sum_{\tau=\underline{t}}^{\bar{t}-h} (MVLPL_{i,\tau+h} - MVLPL_{bcmk,\tau+h})$ , for VARs of different dimensions and across a number of forecast horizons. It is interesting to note that, unlike the conclusions drawn from Figures 1 and 2, there is no strong evidence of a large deterioration in forecasting performance at the time of the financial crisis relative to the univariate benchmark.

The preceding results compared the forecasting performance of various approaches to the AR(1) benchmark. Tables such as Table 5 typically showed strong evidence of statistically significant improvements of all multivariate forecasting methods relative to this benchmark. The online appendix provides additional tables using the BVAR as the benchmark. When using the log predictive likelihoods as the measure of forecast performance, we find that although the compressed VAR approaches do better than the BVAR, this difference is not statistically significant. In fact, we see no significant differences between any of our multivariate forecasting methods. On the other hand, when using MSFEs to evaluate forecast performance we find that in some cases, especially at short horizons, the compressed VARs forecast significantly better than the Minnesota prior BVAR.

We also investigated the robustness of our results to the way we implemented compression in the VARs. While for our main results we focused on two specific approaches, several other ways of performing random compression can be devised, and we experimented with many of those in order to test the reliability and robustness of our findings. As noted previously, a simple way of doing random compression in VARs would be to use the natural conjugate VAR specification in (5), instead of our equation by equation approach in (8). When experimenting with this approach, we found it to forecast very poorly. Other alternative approaches arise from different ways of drawing  $\varphi$ . We considered several alternatives, including the various schemes suggested by Achlioptas (2003), and found that overall these led to very similar results. Next, we tested the robustness of our results to changes in the way the model averaging is done and to the way the variables are ordered in the VAR. Results from both sensitivities are available in the online appendix. As for the first point, we remind the reader that our main results rely on BIC-based weights to perform BMA, with the BICs calculated using the likelihood of the entire  $(n \times 1)$  vector of dependent variables  $Y_t$ . An alternative approach would be to compute the BMA weights by calculating the BICs only relying on the seven variables of interest. In a few cases, altering the BIC weights in this way leads to some improvements, but overall their impact is negligible. As for the second point, in our main results we ordered the variables in  $Y_t$  with our seven variables of interest coming first.

Since our equation-by-equation approach to random compression implies that the different equations will have different right-hand side variables (see (8) and subsequent discussion), it is in principle possible that the way the variables are ordered will matter, especially when we are compressing the error covariance matrix as in  $\text{BCVAR}_c$ . We tested the impact of reorganizing the variables in  $Y_t$  so that the seven variables of interest are ordered last, and found results very similar to those presented here.

Finally, it is worth stressing that this section simply compares the forecast performance of different plausible methods for a particular data set. However, the decision whether to use compression methods should not be based solely on this forecasting comparison. In larger applications, plausible alternatives to random compression such as the Minnesota prior BVAR or any VAR approach that requires the use of MCMC methods may simply be computationally infeasible. In those instances, it may very well be that random compression is the only approach that is computationally feasible.

### 3.5 Forecasting with Time-Variation in Parameters

Before discussing the forecasting results of our compressed TVP-SV-VAR model, it is worthwhile to present some evidence relating to time variation in parameters. For the sake of brevity, we show this for the Medium VAR model only. [Figure 5](#) plots the time series of the predictive density volatilities for the Medium  $\text{BCVAR}_{tvp-sv}$  against the time series of volatilities obtained from the alternative methods described in [section 3](#), and confirms that heteroskedasticity plays a very important role in our data. While the alternative methods allow for some time variation in the volatilities (they are estimated on an expanding window of data),  $\text{BCVAR}_{tvp-sv}$  is finding a lot more variation. This is particularly true around the time of the financial crisis.

[Table 6](#), [Figure 6](#), and [Figure 7](#) present results on the forecast performance of our  $\text{BCVAR}_{tvp-sv}$  approach. The story that jumps out is a strong one: adding time variation in the parameters and volatilities leads to substantial improvements in forecast performance. Conventional wisdom has it that allowing for time-variation (particularly in the error covariance matrix) is particularly important for predictive density estimation. In a time of fluctuating volatility, working with a homoskedastic model may not seriously affect point forecasts, but may lead to poor estimates of higher predictive moments. This wisdom is strongly reinforced by our results. The right panels of [Table 6](#) show that in terms of predictive likelihoods, the  $\text{BCVAR}_{tvp-sv}$  performs much better than our other compressed

VAR approaches, and better (with some exceptions) than standard large VAR and factor methods. This observation is particularly stronger when focusing on the multivariate predictive performance and short to medium forecast horizons. In addition, improvements relative to the univariate benchmark (as indicated by the stars in the table) are almost always strongly statistically significant. In terms of MSFEs, allowing for time variation in parameters leads to some improvements, but these improvements are not as large as those we find when looking at predictive likelihoods. Again, the multivariate results are particularly strong, for all VAR sizes and forecast horizons. In summary, the message conveyed by Table 6 is a particularly strong one:  $\text{BCVAR}_{tvp-sv}$  is forecasting better than any other approach considered in this paper. Figure 6 indicates that, with some exceptions, the reported success in terms of overall point forecast accuracy of the  $\text{BCVAR}_{tvp-sv}$  relative to the alternative methods we considered (namely, DFM, FAVAR, and BVAR) is not the result of any specific and short-lived episode but is instead built gradually throughout the forecast evaluation period, as indicated by the increasing lines depicted in the figure. Interestingly, both at  $h = 1$  and  $h = 12$ , the improvements in forecast performance relative to the various alternatives are particularly notable around the time of the financial crisis, but are not confined to it. Figure 7 provides a similar analysis in terms of the overall density forecast accuracy of the  $\text{BCVAR}_{tvp-sv}$  model. The left panels of the figure show that at  $h = 1$  the previously reported forecast success of the  $\text{BCVAR}_{tvp-sv}$  is once again built steadily throughout the forecast evaluation period. In contrast, the right panels of the figure show that for  $h = 12$ , the 2007-2009 period has a strong negative impact on the density forecast performance of the  $\text{BCVAR}_{tvp-sv}$ .

The preceding results use univariate AR(1) forecasting models as the benchmark for comparison. In the online appendix, we present results using the Minnesota prior BVAR as the benchmark and find that many of the forecast improvements of  $\text{BCVAR}_{tvp-sv}$  over the BVAR are statistically significant. Next, along the lines of our examination of constant coefficient compression models, we investigate the robustness of the  $\text{BCVAR}_{tvp-sv}$  results to changes in the way the model averaging is done and to the way the variables are ordered in the VAR. Results for both sensitivities are reported in the online appendix, where we show that in both cases the forecasting performance of  $\text{BCVAR}_{tvp-sv}$  is hardly affected. There appears to be a slight forecast deterioration when the seven variables of interest are ordered last, but overall our results are quite robust.

We conclude this section with a closer look at the mechanisms through which the

BCVAR<sub>tvp-sv</sub> delivers its superior forecast performance. We start by investigating the relative importance of time-variation in the VAR coefficients versus time-variation in the error covariance matrix. Table 7 compares the forecasting results of the BVAR<sub>tvp-sv</sub> model to those of a model with only variation in the error covariance matrix (BVAR<sub>sv</sub>), and shows that allowing for time variation in both leads to better forecasts. We also wish to compare the performance of the BVAR<sub>tvp-sv</sub> and BVAR<sub>sv</sub> models to other fully Bayesian VAR approaches that allow for time-variation in the parameters. The main difficulty of this exercise is computational. Many of the existing approaches require the use of MCMC methods, which makes them infeasible in large VARs. For instance, the popular TVP-VAR with multivariate stochastic volatility cannot reasonably be scaled up to large VARs due to the computational burden.<sup>17</sup> One recent approach that shows promise for larger VARs with stochastic volatility is that of Carriero, Clark and Marcellino (2016b). Their method improves substantially over the existing algorithms, but still cannot handle the computational demands that come with the very large VARs.<sup>18</sup> Table 7 presents a comparison of their approach (labelled BVAR<sub>ccm</sub> in the table) to ours for the Medium VAR, as well as a Small VAR involving only the seven variables of interest.<sup>19</sup> Results from the three approaches are roughly similar. For the Small VAR, the BVAR<sub>ccm</sub> tends to forecast slightly better at short horizons than the BVAR<sub>sv</sub>. But in the Medium VAR compressed approaches tend to forecast better (particularly when forecast performance is evaluated using WMSFE). Accordingly, we are finding that the BVAR<sub>sv</sub> and BVAR<sub>tvp-sv</sub> models forecast as well or better than a sophisticated fully Bayesian VAR with stochastic volatility where such a comparison is possible. But methods such as Carriero, Clark and Marcellino (2016b), which require the use of MCMC methods, are still not at the stage of being suitable for forecasting with hundreds of variables, much less than thousands of variables that would be possible with random compression.

---

<sup>17</sup>D’Agostino, Gambetti and Giannone (2013) carry out a forecast evaluation exercise using this model with three variables, and even this is very computationally demanding.

<sup>18</sup>Carriero, Clark and Marcellino (2016b) do impulse response analysis in a VAR with 125 variables, but in their forecast evaluation never work with more than 20 variables. Using their model, we produced forecasting results for our Medium VAR. This took 25 hours to run on a PC using a modern Core i7 and 32Gb of RAM.

<sup>19</sup>For the autoregressive coefficients we use the asymmetric Minnesota prior with shrinkage hyperparameter  $\lambda = 0.01$ , and prior mean for own lags  $\delta = 0.95$ . For all other parameters our priors are fairly non-informative and are exactly the same as in Carriero, Clark and Marcellino (2016b).



## 4 Conclusions

In this paper, we draw on ideas from the random projection literature to develop methods suitable for use with large VARs. For such methods to be suitable, they must be computationally simple, theoretically justifiable and empirically successful. We argue that the BCVAR methods developed in this paper meet all these goals. In a substantial macroeconomic application, involving VARs with up to 129 variables, we find BCVAR methods to be fast and yield results which are as good and sometimes better than competing approaches. And, in contrast to the Minnesota prior, random compression methods can easily be scaled up to much higher dimensional VAR models as well as allow for time-variation in the parameters.

## References

- Achlioptas, D. (2003). Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66, 671–687.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70, 191–221.
- Banbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, 25, 71–92.
- Bernanke, B., Boivin, J., Elias, P. (2005). Measuring monetary policy: A factor augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120, 387–422.
- Boot, T. and Nibbering, D. (2016). Forecasting using random subspace methods, Tinbergen Institute Discussion Paper 2016-073/III.
- Carriero, A., Clark, T. and Marcellino, M. (2015). Large vector autoregressions with asymmetric priors and time varying volatilities, manuscript.
- Carriero, A., Clark, T. and Marcellino, M. (2016a). Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics*, forthcoming.
- Carriero, A., Clark, T. and Marcellino, M. (2016b). Large vector autoregressions with stochastic volatility and flexible priors, Federal Reserve Bank of Cleveland Working Paper, no. 16-17.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25, 400–417.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2016). Structural analysis with multivariate autoregressive index models. *Journal of Econometrics* 192, 332–348.

- Chan, J. (2015). Large Bayesian VARs: A flexible Kronecker error covariance structure, manuscript.
- Clark, T. (2011). Real-Time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, 29(3), 327-341.
- Christoffersen, P. and Diebold, F. (1998). Cointegration and long-horizon forecasting. *Journal of Business and Economic Statistics*, 16(4), 450-458.
- D’Agostino, A., Gambetti, L. and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28, 82-101.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253-263.
- Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 1289-1306.
- Doz, C., Giannone, D., Reichlin, L. (2012). A Quasi-maximum likelihood approach for large, approximate dynamic factor models, *Review of Economics and Statistics*, 94, 1014-1024.
- Eisenstat, E., Chan, J. and Strachan, R. (2015). Stochastic model specification search for time-varying parameter VARs, *Econometric Reviews*, forthcoming.
- Elliott, G., Gargano, A. and Timmermann, A. (2013). Complete subset regressions, *Journal of Econometrics*, 177, 357-373.
- Elliott, G., Gargano, A. and Timmermann, A. (2015). Complete subset regressions with large-dimensional sets of predictors, *Journal of Economic Dynamics and Control*, 54, 86-110.
- Gefang, D. (2014). Bayesian doubly adaptive elastic-net lasso for VAR shrinkage. *International Journal of Forecasting*, 30, 1-11.
- George, E., Sun, D. and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142, 553-580.
- Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. J. Aigner and A. S. Goldberger (Eds.), *Latent variables in socio-economic models*, Amsterdam: North-Holland.
- Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75(1), 121-146.
- Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2), 216-230.
- Giannone, D., Lenza, M. and Primiceri, G. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97, 436-451.
- Guhaniyogi, R. and Dunson, D. (2015). Bayesian compressed regression. *Journal of the American Statistical Association*, 110, 1500-1514.
- Harvey, D., Leybourne, S. and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281-291.

- Hoff, P. (2007). Model averaging and dimension selection for the singular value decomposition, *Journal of the American Statistical Association*, 102, 674–685.
- Johnson, W. and Lindenstrauss, J. (1984). Extensions of Lipschitz mapping into Hilbert space. *Contemporary Mathematics*, 26, 189-206.
- Kadiyala, K. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR models. *Journal of Applied Econometrics*, 12, 99-132.
- Kleibergen, F. and Van Dijk, H. (1998). Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory*, 14, 701-743.
- Koop, G. (2008). *Bayesian Econometrics*. Chichester: John Wiley and Sons.
- Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28, 177-203.
- Koop, G. and Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3, 267-358.
- Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177, 185-198.
- Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 28, 204-230.
- McCracken, M. and Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. Federal Reserve Bank of St. Louis, working paper 2015-012A.
- Ng, S. (2016). Opportunities and challenges: Lessons from analyzing terabytes of scanner data. Available at <http://www.columbia.edu/sn2294/papers/sng-worldcongress.pdf>
- Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103, 681-686.
- Park, D., Jun, B. and Kim, J. (1991). Fast tracking RLS algorithm using novel variable forgetting factor with unity zone. *Electronics Letters*, 27, 2150-2151.
- Primiceri, G., (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72, 821-852.
- Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48, 1-48.
- Stock, J. and Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20, 147-162.
- West, K. (1996) Asymptotic inference about predictive ability. *Econometrica*, 64, 1067-1084.
- Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: John Wiley and Sons.

Table 1. Out-of-sample point forecast performance, Medium VAR

Variable	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>
<i>h</i> = 1						<i>h</i> = 2				
PAYEMS	1.082	1.138	0.865	<b>0.830***</b>	0.838***	0.921	1.000	<b>0.554***</b>	0.728***	0.732***
CPIAUCSL	1.142	1.017	<b>0.949</b>	0.958	0.967	1.086	1.037	0.999	0.940	<b>0.936*</b>
FEDFUNDS	2.278	1.848	2.760	1.023	<b>0.962</b>	1.441	1.424	2.448	0.974	<b>0.945</b>
INDPRO	0.863***	0.879**	<b>0.810**</b>	0.828***	0.889***	0.909	0.952	<b>0.825*</b>	0.931	0.929*
UNRATE	0.878	0.840**	<b>0.783***</b>	0.803***	0.848***	0.894	0.908	<b>0.805**</b>	0.844***	0.869**
PPIFGS	1.000	1.002	0.980	<b>0.970</b>	0.993	1.052	1.037	1.083	1.029	<b>1.012</b>
GS10	1.141	<b>0.988</b>	1.092	0.996	1.013	1.038	1.023	1.082	<b>1.003</b>	1.003
<i>h</i> = 3						<i>h</i> = 6				
PAYEMS	0.846	0.915	<b>0.522***</b>	0.683***	0.687***	0.951	0.903	<b>0.686*</b>	0.747**	0.738**
CPIAUCSL	1.096	1.031	1.042	0.982	<b>0.978</b>	1.042	<b>0.979</b>	1.057	1.003	0.995
FEDFUNDS	1.289	1.272	1.858	1.017	<b>1.001</b>	1.198	1.017	1.195	0.991	<b>0.986</b>
INDPRO	<b>0.928</b>	0.991	0.931	0.939	0.949	0.959	1.024	1.024	0.970	<b>0.957</b>
UNRATE	0.942	0.959	<b>0.850*</b>	0.871**	0.866***	0.993	0.995	0.947	<b>0.939*</b>	0.946*
PPIFGS	1.032	<b>1.016</b>	1.102	1.050	1.042	1.047	<b>1.026</b>	1.135	1.059	1.043
GS10	1.038	1.036	1.140	1.046	<b>1.032</b>	<b>1.006</b>	1.015	1.115	1.036	1.038
<i>h</i> = 9						<i>h</i> = 12				
PAYEMS	1.005	0.936	<b>0.824</b>	0.838	0.843	1.015	0.963	<b>0.931</b>	0.934	0.935
CPIAUCSL	1.001	<b>0.960</b>	1.036	0.979	0.961	1.007	<b>0.969</b>	1.069	1.016	1.012
FEDFUNDS	1.133	0.945	0.991	<b>0.921</b>	0.950	1.137	<b>0.975</b>	1.077	0.991	0.996
INDPRO	<b>0.958</b>	1.009	1.024	0.967	0.978	0.981	1.011	1.004	<b>0.974</b>	0.975
UNRATE	1.009	1.001	0.972	0.954	<b>0.951</b>	1.007	1.010	1.008	<b>0.968</b>	0.968
PPIFGS	1.017	<b>1.004</b>	1.116	1.055	1.042	1.018	<b>1.000</b>	1.140	1.070	1.053
GS10	<b>0.997</b>	0.997	1.025	1.005	1.016	1.012	<b>1.000</b>	1.052	1.029	1.023

This table reports the ratio between the MSFE of model  $i$  and the MSFE of the benchmark AR(1) for the Medium VAR, computed as  $MSFE_{ijh} = \sum_{\tau=\underline{t}}^{\bar{t}-h} e_{i,j,\tau+h}^2 / \sum_{\tau=\underline{t}}^{\bar{t}-h} e_{bckm,j,\tau+h}^2$ , where  $e_{i,j,\tau+h}^2$  and  $e_{bckm,j,\tau+h}^2$  are the squared forecast errors of variable  $j$  at time  $\tau$  and forecast horizon  $h$  generated by model  $i$  and the AR(1) model, respectively while  $\underline{t}$  and  $\bar{t}$  denote the start and end of the out-of-sample period. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. \* significance at the 10% level; \*\* significance at the 5% level; \*\*\* significance at the 1% level.

Table 2. Out-of-sample density forecast performance, Medium VAR

Variable	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>
<i>h</i> = 1						<i>h</i> = 2				
PAYEMS	0.066***	0.030	<b>0.218***</b>	0.086***	0.083***	0.117***	0.061*	<b>0.366***</b>	0.158***	0.163***
CPIAUCSL	-0.115	-0.055	-0.674	0.003	<b>0.156</b>	-0.266	-0.280	-1.669	-0.263	<b>-0.247</b>
FEDFUNDS	-0.012	0.043***	<b>0.131***</b>	0.006	0.005	0.028	0.042***	<b>0.115**</b>	0.022***	0.022***
INDPRO	-0.105	<b>0.046</b>	-0.098	-0.063	0.028	0.008	0.028	-0.049	0.084**	<b>0.109**</b>
UNRATE	0.083**	0.121***	<b>0.167***</b>	0.105***	0.081***	0.072**	0.060**	<b>0.131***</b>	0.077***	0.062***
PPIFGS	<b>0.025</b>	-0.033	-0.448	-0.071	0.020	-0.043	-0.135	-0.725	<b>0.019</b>	-0.063
GS10	-0.029	0.007	<b>0.015</b>	-0.001	-0.007	-0.011	-0.017	-0.009	<b>-0.008</b>	-0.016
<i>h</i> = 3						<i>h</i> = 6				
PAYEMS	0.124***	0.085**	<b>0.364***</b>	0.172***	0.185***	0.050	0.071	<b>0.245***</b>	0.144***	0.168***
CPIAUCSL	0.034	<b>0.043</b>	-0.984	-0.095	-0.017	-0.007	<b>0.004</b>	-0.860	-0.220	-0.249
FEDFUNDS	0.021	0.023*	<b>0.115***</b>	0.014	0.014*	0.013	0.015**	<b>0.119***</b>	0.017**	0.011
INDPRO	<b>0.144</b>	0.090	-0.001	0.125	0.073***	-0.005	<b>0.052</b>	-0.227	-0.014	0.038***
UNRATE	0.041	0.024	<b>0.109***</b>	0.065***	0.062***	0.022	0.007	<b>0.058***</b>	0.042***	0.040***
PPIFGS	-0.081	0.044	-0.483	<b>0.049</b>	-0.098	-0.063	<b>0.003</b>	-0.807	-0.172	-0.100
GS10	0.012	<b>0.014</b>	0.010	0.013	0.003	<b>0.003</b>	0.001	0.002	-0.003	-0.013
<i>h</i> = 9						<i>h</i> = 12				
PAYEMS	-0.005	0.038	0.092	<b>0.096***</b>	0.084***	0.023	0.038	0.040	0.074***	<b>0.089***</b>
CPIAUCSL	-0.022	<b>0.220</b>	-0.746	-0.083	-0.184	-0.091	<b>-0.037</b>	-0.905	-0.254	-0.312
FEDFUNDS	0.007	0.008	<b>0.119***</b>	0.008	0.005	-0.014	-0.002	<b>0.109***</b>	-0.006	-0.008
INDPRO	-0.038	-0.067	-0.152	<b>-0.012</b>	-0.077	0.098	-0.007	-0.018	0.128	<b>0.149</b>
UNRATE	0.015	0.010	0.040	<b>0.048***</b>	0.036***	-0.002	0.000	<b>0.033</b>	0.024**	0.020**
PPIFGS	-0.006	<b>0.106</b>	-0.413	-0.070	0.060	-0.001	<b>0.120</b>	-0.391	-0.144	-0.108
GS10	0.009	0.009**	<b>0.041**</b>	0.011	0.001	-0.016	-0.001	<b>0.010</b>	-0.003	-0.014

This table reports the average log predictive likelihood (ALPL) differential between model  $i$  and the benchmark AR(1) for the Medium VAR, computed as  $ALPL_{ijh} = \sum_{\tau=\underline{t}}^{\bar{t}-h} (LPL_{i,j,\tau+h} - LPL_{bckm,j,\tau+h}) / (\bar{t} - \underline{t} - h + 1)$ , where  $LPL_{i,j,\tau+h}$  and  $LPL_{bckm,j,\tau+h}$  are the log predictive likelihoods of variable  $j$  at time  $\tau$  and forecast horizon  $h$  generated by model  $i$  and the AR(1) model, respectively while  $\underline{t}$  and  $\bar{t}$  denote the start and end of the out-of-sample period. All density forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate the highest ALPL across all models for a given variable-forecast horizon pair. \* significance at the 10% level; \*\* significance at the 5% level; \*\*\* significance at the 1% level.

Table 3. Out-of-sample point forecast performance, Large VAR

Variable	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>
<i>h</i> = 1						<i>h</i> = 2				
PAYEMS	0.789**	1.068	<b>0.748***</b>	0.777***	0.796***	0.710*	0.801	<b>0.481***</b>	0.640***	0.671***
CPIAUCSL	0.930	0.925	<b>0.860**</b>	0.928**	0.935*	1.003	0.996	0.932	<b>0.887**</b>	0.892**
FEDFUNDS	2.120	1.669	2.061	<b>0.965</b>	1.013	1.766	1.338	2.178	0.962	<b>0.892</b>
INDPRO	0.830**	0.858**	<b>0.778***</b>	0.844***	0.902***	0.860	0.884	<b>0.801*</b>	0.945	0.920**
UNRATE	0.807**	<b>0.740***</b>	0.796**	0.810***	0.860***	0.811**	0.829**	<b>0.769**</b>	0.852***	0.852***
PPIFGS	0.940	0.984	<b>0.938</b>	0.974	1.012	1.065	1.047	1.063	<b>1.013</b>	1.019
GS10	1.111	1.037	1.103	<b>1.009</b>	1.015	1.036	1.057	1.136	<b>1.005</b>	1.044
<i>h</i> = 3						<i>h</i> = 6				
PAYEMS	0.715	0.726	<b>0.474***</b>	0.611***	0.622***	0.923	0.828	<b>0.620</b>	0.668**	0.706**
CPIAUCSL	0.979	0.988	0.979	0.912	<b>0.904*</b>	0.961	0.922	1.044	0.931	<b>0.916</b>
FEDFUNDS	1.526	1.104	1.819	<b>0.967</b>	0.987	1.395	<b>0.959</b>	1.325	0.991	0.988
INDPRO	0.943	0.950	<b>0.893</b>	0.950	0.938	1.035	0.977	1.022	<b>0.967</b>	0.983
UNRATE	0.888	0.868*	<b>0.836*</b>	0.876**	0.882***	0.981	0.931*	<b>0.886*</b>	0.924**	0.943*
PPIFGS	1.086	1.040	1.089	<b>1.034</b>	1.048	1.112	1.057	1.151	1.063	<b>1.041</b>
GS10	1.067	1.094	1.215	<b>1.049</b>	1.064	1.073	1.038	1.179	<b>1.022</b>	1.042
<i>h</i> = 9						<i>h</i> = 12				
PAYEMS	1.001	0.916	<b>0.743</b>	0.766	0.760*	1.065	0.996	0.870	<b>0.848</b>	0.866
CPIAUCSL	0.944	0.887**	1.022	0.895	<b>0.885</b>	0.947	0.915***	1.036	0.901	<b>0.872**</b>
FEDFUNDS	1.279	0.995	1.115	<b>0.969</b>	0.995	1.225	<b>0.976</b>	1.151	1.023	1.035
INDPRO	1.043	1.004	1.068	<b>0.975</b>	0.990	0.993	0.997	1.074	<b>0.989</b>	1.012
UNRATE	1.019	0.967*	<b>0.938</b>	0.951	0.957	1.014	0.981	0.982	<b>0.979</b>	0.989
PPIFGS	1.060	<b>1.011</b>	1.149	1.047	1.035	1.100	<b>1.032</b>	1.182	1.073	1.042
GS10	1.023	<b>1.000</b>	1.074	1.006	1.024	1.034	<b>1.003</b>	1.081	1.013	1.006

This table reports the ratio between the MSFE of model  $i$  and the MSFE of the benchmark AR(1) for the Large VAR, across a number of different forecast horizons  $h$ . See notes under Table 1 for additional details.

Table 4. Out-of-sample density forecast performance, Large VAR

Variable	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>	DFM	FAVAR	BVAR	BCVAR	BCVAR <sub>c</sub>
<i>h</i> = 1						<i>h</i> = 2				
PAYEMS	0.189***	0.061***	<b>0.302***</b>	0.104***	0.102***	0.224***	0.155***	<b>0.471***</b>	0.196***	0.196***
CPIAUCSL	-0.005	0.041	-0.362	0.025	<b>0.052</b>	-0.419	-0.210	-2.118	<b>0.098***</b>	0.095**
FEDFUNDS	0.030	0.052***	<b>0.291***</b>	0.014**	0.010**	0.019	0.036*	<b>0.247***</b>	0.013*	0.014**
INDPRO	-0.051	-0.029	-0.311	<b>0.092***</b>	0.026	<b>0.238*</b>	0.170**	-0.057	0.041	0.179
UNRATE	0.130***	<b>0.157***</b>	0.125**	0.095***	0.079***	0.102***	0.092***	<b>0.163***</b>	0.076***	0.079***
PPIFGS	-0.111	0.002	-1.029	<b>0.059*</b>	-0.087	-0.241	-0.157	-1.813	-0.064	<b>-0.015</b>
GS10	-0.008	-0.007	<b>0.006</b>	-0.001	0.000	0.006	-0.010	-0.009	<b>0.012</b>	-0.001
<i>h</i> = 3						<i>h</i> = 6				
PAYEMS	0.197***	0.168***	<b>0.447***</b>	0.229***	0.225***	0.090*	0.097**	<b>0.296***</b>	0.199***	0.191***
CPIAUCSL	-0.190	-0.070	-2.294	0.000	<b>0.121***</b>	-0.119	0.087	-2.185	<b>0.227</b>	0.042
FEDFUNDS	0.016	0.032*	<b>0.228***</b>	0.022**	0.016**	0.003	0.013*	<b>0.186***</b>	0.007	0.013
INDPRO	-0.025	0.029	<b>0.065</b>	0.052***	0.043***	<b>0.082</b>	-0.028	-0.151	0.056*	-0.088
UNRATE	0.059**	0.061***	<b>0.106**</b>	0.067***	0.048***	0.017	0.028**	<b>0.084***</b>	0.036**	0.030***
PPIFGS	-0.283	-0.002	-1.315	<b>0.086</b>	-0.062	-0.124	-0.100	-1.594	<b>0.003</b>	-0.173
GS10	0.018	0.012	-0.027	<b>0.032</b>	0.009	-0.014	0.000	-0.024	<b>0.012</b>	-0.005
<i>h</i> = 9						<i>h</i> = 12				
PAYEMS	0.005	0.037	0.128	<b>0.129***</b>	0.123***	0.019	0.019	0.077	0.100***	<b>0.110***</b>
CPIAUCSL	<b>0.212</b>	0.002	-0.995	-0.032	0.059	<b>0.060</b>	-0.239	-1.661	0.016	-0.171
FEDFUNDS	0.004	0.011***	<b>0.275***</b>	0.014*	0.010	0.002	0.007	<b>0.211***</b>	-0.002	-0.001
INDPRO	<b>0.110</b>	0.011	-0.183	0.081	0.050**	<b>0.062</b>	-0.038	-0.174	0.021*	-0.057
UNRATE	-0.002	0.007	<b>0.045</b>	0.026*	0.028**	0.008	0.019**	<b>0.034</b>	0.029**	0.021**
PPIFGS	0.022	0.064	-1.227	<b>0.099</b>	0.039	-0.189	<b>-0.130</b>	-0.724	-0.144	-0.274
GS10	-0.003	0.017	<b>0.039</b>	0.008	-0.011	-0.002	0.007	<b>0.034</b>	-0.005	-0.021

This table reports the average log predictive likelihood (ALPL) differential between model  $i$  and the benchmark AR(1) for the Large VAR, across a number of different forecast horizons  $h$ . See notes under Table 2 for additional details.

Table 5. Out-of-sample forecast performance: Multivariate results

$Fcst\ h.$	<i>Medium VAR</i>									
	<i>WMSFE</i>					<i>MVALPL</i>				
	<i>DFM</i>	<i>FAVAR</i>	<i>BVAR</i>	<i>BCVAR</i>	<i>BCVAR<sub>c</sub></i>	<i>DFM</i>	<i>FAVAR</i>	<i>BVAR</i>	<i>BCVAR</i>	<i>BCVAR<sub>c</sub></i>
h= 1	1.158	1.066	1.132	<b>0.916***</b>	0.935***	0.551***	0.770***	<b>0.979***</b>	0.925***	0.285***
h= 2	1.051	1.052	1.115	0.929**	<b>0.926***</b>	0.832***	0.818***	<b>1.068***</b>	1.021***	0.401***
h= 3	1.027	1.031	1.064	0.944*	<b>0.940*</b>	0.890***	0.874***	<b>1.097***</b>	1.046***	0.356***
h= 6	1.027	0.992	1.017	0.961	<b>0.954</b>	0.868***	0.837***	<b>1.030***</b>	1.009***	0.296***
h= 9	1.017	0.977	0.995	<b>0.957</b>	0.960	0.850***	0.858***	<b>1.021***</b>	1.017***	0.254***
h=12	1.025	<b>0.988</b>	1.039	0.996	0.994	0.877***	0.867***	<b>0.927***</b>	0.886***	0.176***
	<i>Large VAR</i>									
	<i>DFM</i>	<i>FAVAR</i>	<i>BVAR</i>	<i>BCVAR</i>	<i>BCVAR<sub>c</sub></i>	<i>DFM</i>	<i>FAVAR</i>	<i>BVAR</i>	<i>BCVAR</i>	<i>BCVAR<sub>c</sub></i>
h= 1	1.049	1.009	1.017	<b>0.907***</b>	0.940***	0.950***	0.935***	0.905***	<b>0.996***</b>	0.303***
h= 2	1.037	0.996	1.053	0.909***	<b>0.908***</b>	1.053***	0.971***	0.944***	<b>1.139***</b>	0.406***
h= 3	1.030	0.970	1.045	<b>0.916**</b>	0.922**	1.049***	0.999***	0.974***	<b>1.179***</b>	0.368***
h= 6	1.063	0.955	1.026	<b>0.933</b>	0.940	0.957***	0.995***	0.830***	<b>1.131***</b>	0.269***
h= 9	1.049	0.965	1.009	<b>0.938</b>	0.943	0.972***	0.954***	0.879***	<b>1.076***</b>	0.243***
h=12	1.052	0.984	1.049	0.969	<b>0.968</b>	0.934***	0.910***	0.709**	<b>1.009***</b>	0.145

The left half of this table reports the ratio between the multivariate weighted mean squared forecast error (WMSFE) of model  $i$  and the WMSFE of the benchmark AR(1) model, computed as

$$WMSFE_{ih} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{i,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{bcmk,\tau+h}},$$

where  $we_{i,\tau+h} = (e'_{i,\tau+h} \times W \times e_{i,\tau+h})$  and  $we_{bcmk,\tau+h} = (e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h})$  denote the weighted forecast errors of model  $i$  and the benchmark model at time  $\tau + h$ ,  $e_{i,\tau+h}$  and  $e_{bcmk,\tau+h}$  are the  $(N \times 1)$  vector of forecast errors, and  $W$  is an  $(N \times N)$  matrix of weights. We set  $N = 7$ , to focus on the following key seven series,  $\{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$ . In addition, we set the matrix  $W$  to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast.  $\underline{t}$  and  $\bar{t}$  denote the start and end of the out-of-sample period,  $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ , and  $h \in \{1, 2, 3, 6, 9, 12\}$ . The right half of the table shows the multivariate average log predictive likelihood differentials between model  $i$  and the benchmark AR(1), computed as

$$MVALPL_{ih} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} (MVLPL_{i,\tau+h} - MVLPL_{bcmk,\tau+h}),$$

where  $MVLPL_{i,\tau+h}$  and  $MVLPL_{bcmk,\tau+h}$  denote the multivariate log predictive likelihoods of model  $i$  and the benchmark model at time  $\tau + h$ , and are computed under the assumption of joint normality. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate the lowest WMSFE and highest MVALPL across all models for any given VAR size - forecast horizon pair. \* significance at the 10% level; \*\* significance at the 5% level; \*\*\* significance at the 1% level.

Table 6. Out-of-sample forecast performance: Compressed TVP-SV VAR

Variable	Medium VAR											
	MSFE						ALPL					
	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
PAYEMS	<b>0.700***</b>	0.565***	0.565***	<b>0.651**</b>	<b>0.769*</b>	<b>0.872</b>	<b>0.338***</b>	<b>0.391***</b>	0.352***	0.078	-0.422	-0.533
CPIAUCSL	<b>0.924**</b>	<b>0.872***</b>	<b>0.884***</b>	<b>0.869**</b>	<b>0.841***</b>	<b>0.845***</b>	<b>0.284*</b>	<b>0.211***</b>	<b>0.461</b>	<b>0.191</b>	<b>0.280</b>	<b>0.292</b>
FEDFUNDS	<b>0.879*</b>	<b>0.892</b>	<b>0.924</b>	0.995	0.967	1.061	<b>0.760***</b>	<b>0.594**</b>	<b>0.423</b>	<b>0.382</b>	<b>0.303</b>	<b>0.365</b>
INDPRO	0.899***	0.925*	0.940	0.978	0.980	0.989	-0.030	-0.224	-0.128	-0.509	-0.414	-0.255
UNRATE	0.846***	0.847**	0.876*	0.939	0.971	1.011	0.123***	0.104***	0.095***	<b>0.059***</b>	0.036	-0.009
PPIFGS	<b>0.968</b>	<b>0.991</b>	<b>1.001</b>	<b>0.998</b>	<b>0.992</b>	1.010	<b>0.270*</b>	<b>0.349</b>	<b>0.401</b>	<b>0.283</b>	<b>0.407</b>	<b>0.354</b>
GS10	1.018	1.017	1.039	1.030	<b>0.995</b>	1.030	<b>0.025</b>	-0.016	-0.053	-0.057	-0.004	<b>0.030</b>
Multivariate	<b>0.905***</b>	<b>0.884***</b>	<b>0.892***</b>	<b>0.916*</b>	<b>0.924*</b>	<b>0.967</b>	<b>1.653***</b>	<b>1.701***</b>	<b>1.573***</b>	<b>1.224***</b>	<b>1.049***</b>	0.851***
	Large VAR											
	MSFE						ALPL					
	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
PAYEMS	<b>0.685***</b>	0.566***	0.548***	0.656*	0.762	0.879	<b>0.338***</b>	0.405***	0.374***	0.083	-0.447	-0.530
CPIAUCSL	0.904**	<b>0.846***</b>	<b>0.844***</b>	<b>0.848**</b>	<b>0.800***</b>	<b>0.796***</b>	<b>0.241</b>	<b>0.364*</b>	<b>0.361</b>	<b>0.354</b>	<b>0.539</b>	<b>0.074</b>
FEDFUNDS	<b>0.885</b>	0.911	<b>0.920</b>	1.022	1.034	1.075	<b>0.715***</b>	<b>0.577*</b>	<b>0.489</b>	<b>0.445</b>	0.100	<b>0.269</b>
INDPRO	0.896***	0.928	0.957	0.996	1.002	1.020	<b>0.116**</b>	0.036	-0.184	-0.320	-0.205	-0.210
UNRATE	0.836***	0.851**	0.880*	0.949	0.981	1.026	0.122***	0.102***	0.078***	0.050**	0.034	0.010
PPIFGS	0.983	<b>0.985</b>	<b>1.005</b>	<b>1.008</b>	<b>0.995</b>	<b>1.012</b>	<b>0.254*</b>	<b>0.363</b>	<b>0.371</b>	<b>0.346</b>	<b>0.385</b>	<b>0.213</b>
GS10	1.021	1.021	<b>1.034</b>	1.024	1.013	1.021	<b>0.008</b>	<b>0.037**</b>	0.017	0.008	0.029	-0.033
Multivariate	<b>0.902***</b>	<b>0.883***</b>	<b>0.885**</b>	<b>0.922</b>	<b>0.932</b>	<b>0.967</b>	<b>1.667***</b>	<b>1.666***</b>	<b>1.593***</b>	<b>1.216***</b>	1.002***	0.713*

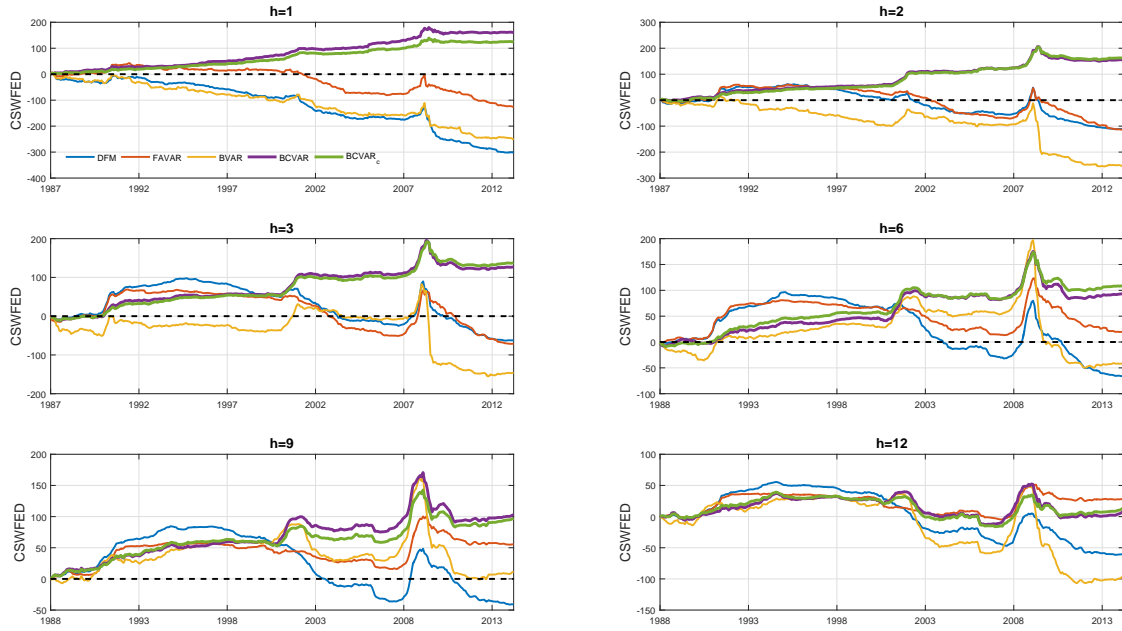
The left half of this table reports the ratio between the univariate or multivariate weighted mean squared forecast error of the  $BCVAR_{tvp-sv}$  model and the univariate or multivariate weighted mean squared forecast error of the benchmark AR(1) model. The right half of the table shows the univariate or multivariate average log predictive likelihood differentials between the  $BCVAR_{tvp-sv}$  model and the benchmark AR(1) model.  $h$  denotes the forecast horizons, with  $h \in \{1, 2, 3, 6, 9, 12\}$ . All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate all instances where the  $BCVAR_{tvp-sv}$  model outperforms all alternative models for any given VAR size/variable/forecast horizon combination. \* significance at the 10% level; \*\* significance at the 5% level; \*\*\* significance at the 1% level.

Table 7. Out-of-sample forecast performance: Multivariate results, alternative SV models

Fcst $h$ .	Small VAR					
	WMSFE			MVALPL		
	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$
$h=1$	<b>0.917***</b>	0.942***	0.918***	<b>2.047***</b>	1.696***	1.719***
$h=2$	0.930***	0.944***	<b>0.895***</b>	<b>1.907***</b>	1.654***	1.745***
$h=3$	0.936***	0.951**	<b>0.901***</b>	<b>1.845***</b>	1.563***	1.645***
$h=6$	0.946***	0.971	<b>0.912***</b>	<b>1.608***</b>	1.228***	1.386***
$h=9$	0.968***	0.981	<b>0.936***</b>	<b>1.385***</b>	0.978***	1.143***
$h=12$	0.992	0.999	<b>0.960*</b>	<b>0.931*</b>	0.811*	0.930***
	Medium VAR					
	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$
	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$
$h=1$	1.070	0.935***	<b>0.905***</b>	1.599***	1.522***	<b>1.653***</b>
$h=2$	1.089	0.922***	<b>0.884***</b>	1.521***	1.558***	<b>1.701***</b>
$h=3$	1.123	0.931**	<b>0.892***</b>	1.236***	1.399***	<b>1.573***</b>
$h=6$	1.125	0.937*	<b>0.916*</b>	1.041***	1.129***	<b>1.224***</b>
$h=9$	1.031	0.947*	<b>0.924*</b>	<b>1.078***</b>	0.938***	1.049***
$h=12$	1.007	0.981	<b>0.967</b>	<b>1.039***</b>	0.760**	0.851***
	Large VAR					
	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$
	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$	$BVAR_{ccm}$	$BCVAR_{sv}$	$BCVAR_{tvp-sv}$
$h=1$		0.942***	<b>0.902***</b>		1.488***	<b>1.667***</b>
$h=2$		0.924**	<b>0.883***</b>		1.543***	<b>1.666***</b>
$h=3$		0.919**	<b>0.885**</b>		1.394***	<b>1.593***</b>
$h=6$		0.939	<b>0.922</b>		1.118***	<b>1.216***</b>
$h=9$		0.938	<b>0.932</b>		0.894***	<b>1.002***</b>
$h=12$		<b>0.950</b>	0.967		<b>0.722*</b>	0.713*

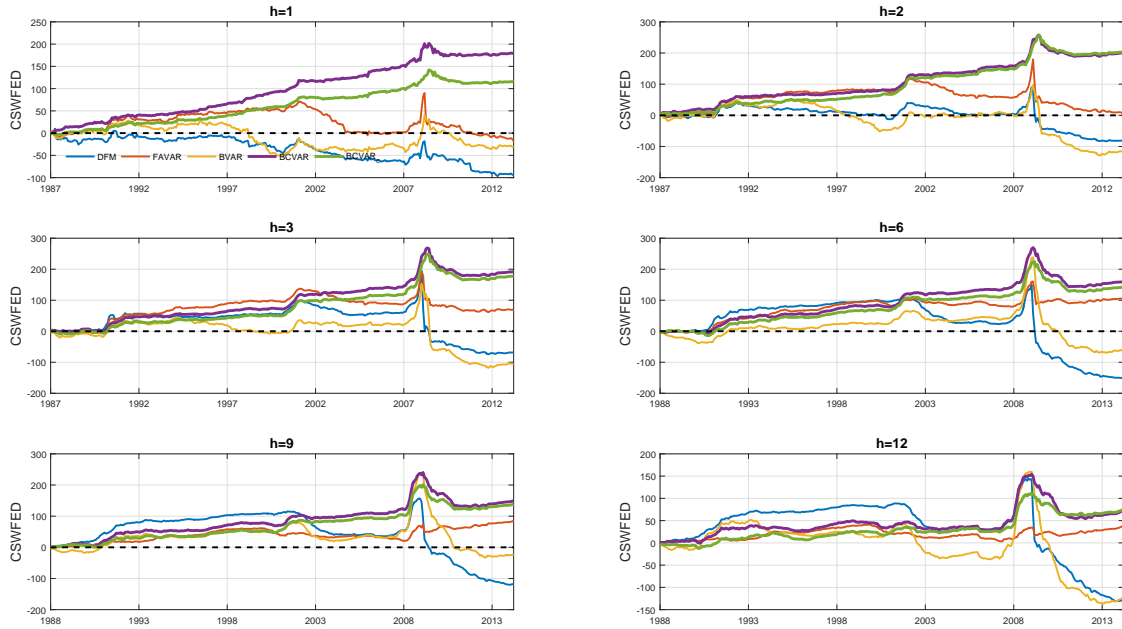
The left half of this table reports the ratio between the multivariate weighted mean squared forecast error (WMSFE) of model  $i$  and the WMSFE of the benchmark AR(1) model for different forecast horizons  $h$  and VAR size, where  $i \in \{BVAR_{ccm}, BCVAR_{sv}, BCVAR_{tvp-sv}\}$ , and  $h \in \{1, 2, 3, 6, 9, 12\}$ . The right half of the table shows the multivariate average log predictive likelihood differentials between model  $i$  and the benchmark AR(1). See notes to Table 5 for additional details.

Figure 1. Cumulative sum of weighted forecast error differentials, Medium VAR



This figure plots the cumulative sum of weighted forecast errors generated by the AR(1) model minus the cumulative sum of weighted forecast errors generated by model  $i$  for a Medium VAR and forecast horizon  $h$ , where  $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ , and  $h \in \{1, 2, 3, 6, 9, 12\}$ . All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Each panel displays results for a different forecast horizon.

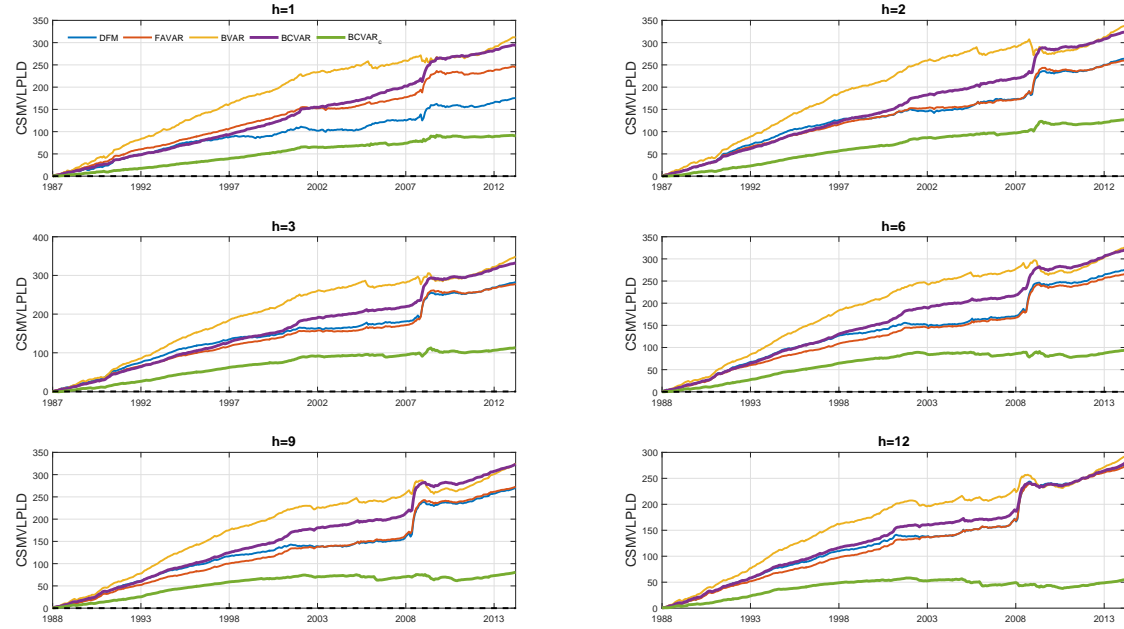
Figure 2. Cumulative sum of weighted forecast error differentials, Large VAR



This figure plots the cumulative sum of weighted forecast errors generated by the AR(1) model minus the cumulative sum of weighted forecast errors generated by model  $i$  for a Large VAR.  $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ . See notes to [Figure 1](#) for additional details.

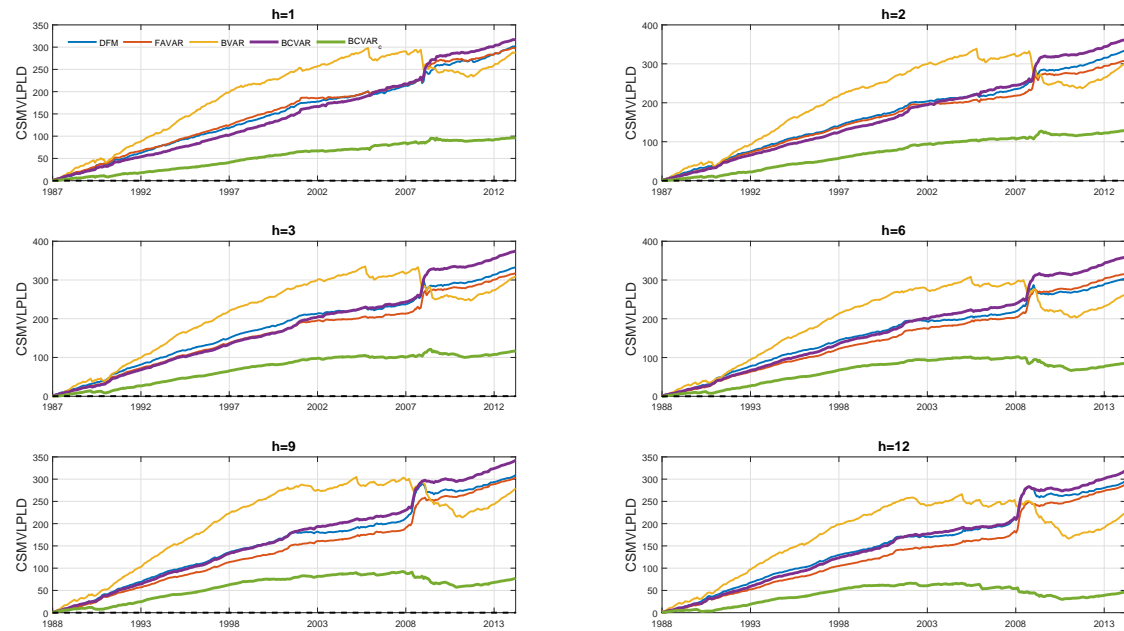


Figure 3. Cumulative sum of multivariate log predictive likelihood differentials, Medium VAR



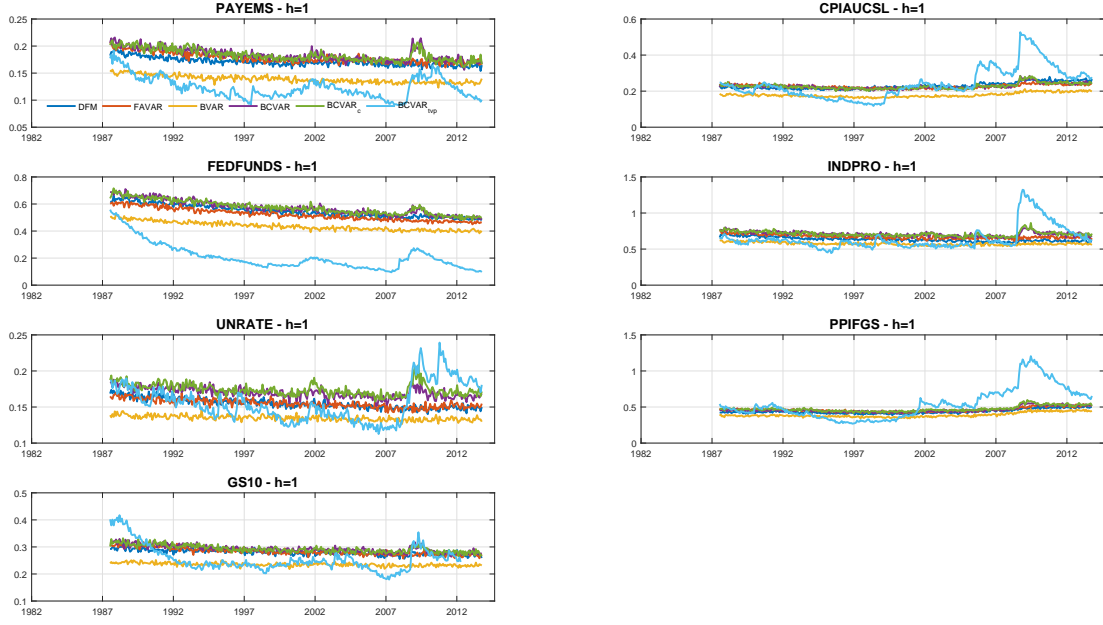
This figure plots the cumulative sum of the multivariate log predictive likelihoods generated by model  $i$  minus the cumulative sum of the multivariate log predictive likelihoods computed from an AR(1) model for a Medium VAR, where  $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ , and  $h \in \{1, 2, 3, 6, 9, 12\}$ . All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Each panel displays results for a different forecast horizon.

Figure 4. Cumulative sum of multivariate log predictive likelihood differentials, Large VAR



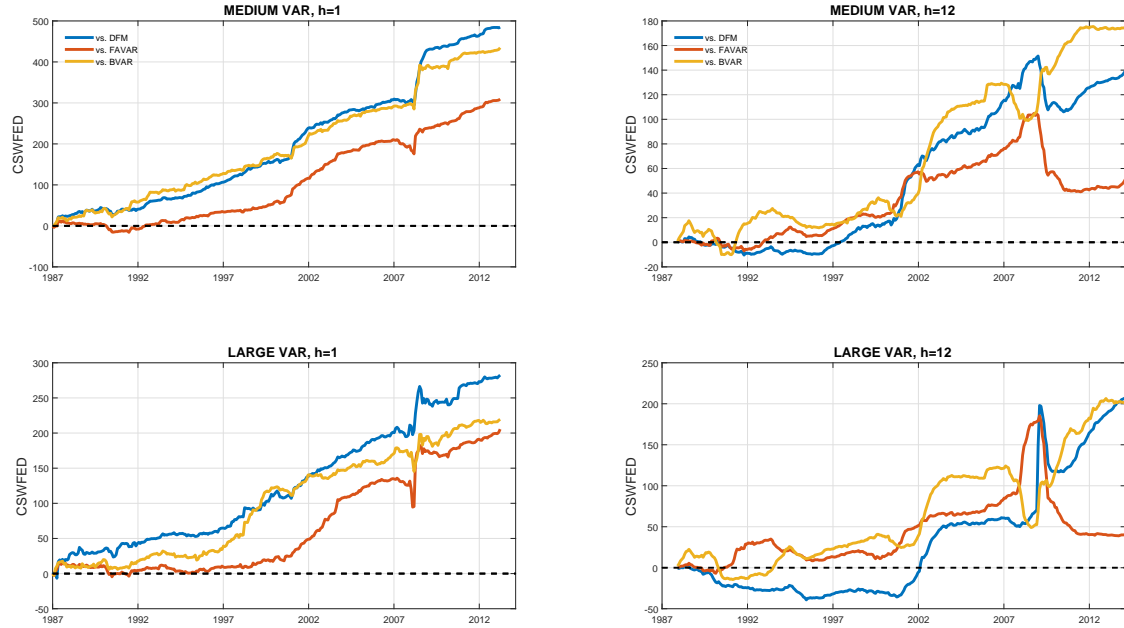
This figure plots the cumulative sum of the multivariate log predictive likelihoods generated by model  $i$  minus the cumulative sum of the multivariate log predictive likelihoods computed from an AR(1) model for a Large VAR.  $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ . See notes to Figure 3 for additional details.

Figure 5. Predictive density volatilities, Medium VAR



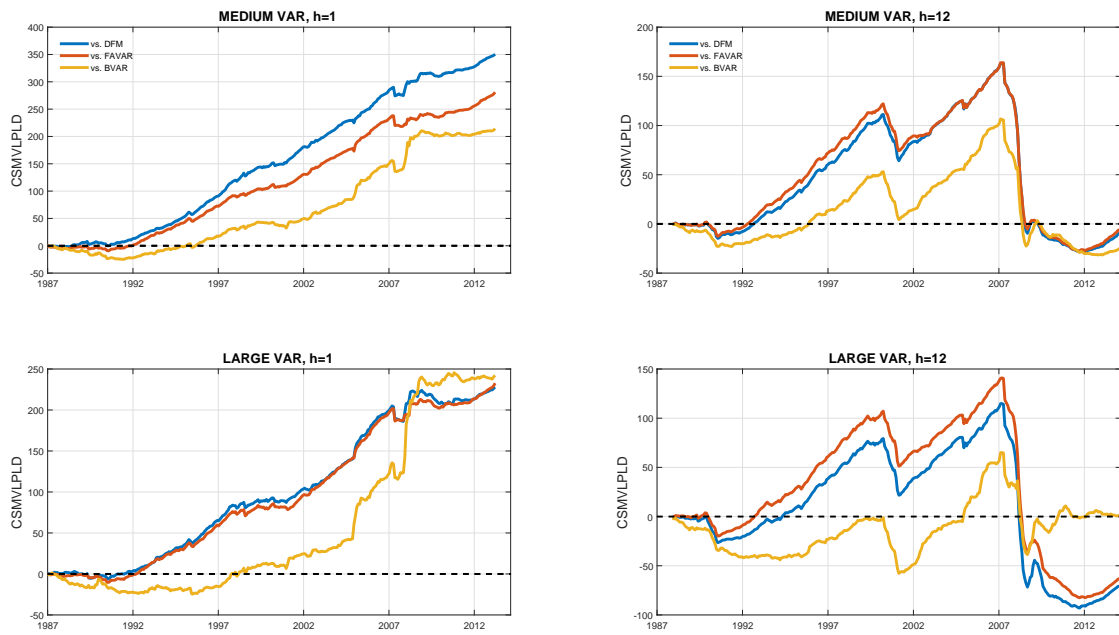
This figure plots the time series of the predicted volatilities over the entire out-of-sample period, for  $h = 1$  and the different models entertained,  $\{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp-sv}\}$ . The out of sample period starts in 1987:07 and ends in 2014:12. Each panel displays results for a different variable  $j$ , where  $j \in \{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$ .

Figure 6. Cumulative sum of weighted forecast error differentials, Compressed TVP-SV VAR



This figure plots the cumulative sum of weighted forecast errors generated by either the DFM, FAVAR, or BVAR models minus the cumulative sum of weighted forecast errors generated by the  $BCVAR_{tvp-sv}$  model for different VAR sizes and forecast horizons. See notes to Table 5 for additional details. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12.

Figure 7. Cumulative sum of multivariate log predictive likelihood differentials, Compressed TVP-SV VAR



This figure plots the cumulative sum of the multivariate log predictive likelihoods generated by the  $BCVAR_{tvp-sv}$  model minus the cumulative sum of the multivariate log predictive likelihoods computed from either the DFM, FAVAR, or BVAR model for different VAR sizes and forecast horizons. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12.